



ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΜΗΜΑΤΟΣ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΣΧΕΔΙΑΣΗ ΚΑΙ ΑΝΑΠΤΥΞΗ ΔΙΑΔΡΑΣΤΙΚΗΣ ΔΙΕΠΑΦΗΣ ΧΡΗΣΤΗ ΓΙΑ ΤΗΝ
ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΕΞΟΡΥΞΗΣ ΓΝΩΜΗΣ

Διπλωματική Εργασία

του

Μουσενίκα Δημήτριου

Θεσσαλονίκη, Φεβρουάριος 2021

ΣΧΕΔΙΑΣΗ ΚΑΙ ΑΝΑΠΤΥΞΗ ΔΙΑΔΡΑΣΤΙΚΗΣ ΔΙΕΠΑΦΗΣ ΧΡΗΣΤΗ ΓΙΑ ΤΗΝ
ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΕΞΟΡΥΞΗΣ ΓΝΩΜΗΣ

Μουσενίκας Δημήτριος

Δίπλωμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, ΑΠΘ, 2015

Διπλωματική Εργασία

υποβαλλόμενη για τη μερική εκπλήρωση των απαιτήσεων του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΤΙΤΛΟΥ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΠΛΗΡΟΦΟΡΙΚΗ

Επιβλέπουσα Καθηγήτρια
Κολωνιάρη Γεωργία

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26/02/2021

Κολωνιάρη Γεωργία

Ξυνόγαλος Στυλιανός

Ψάννης Κωνσταντίνος

.....

.....

.....

Μουσενίκας Δημήτριος

.....

Περίληψη

Η παρούσα εργασία πραγματεύεται το θέμα της ανάλυσης συναισθημάτων και ειδικότερα το πώς μπορούν να παρουσιαστούν τα αποτελέσματα μια τέτοιας ανάλυσης, ώστε να μπορούν να γίνουν ευκολότερα κατανοητά και να αξιοποιηθούν με αποτελεσματικότερο τρόπο. Στο πρώτο τμήμα της εργασίας έχει πραγματοποιηθεί ευρεία βιβλιογραφική ανασκόπηση του θεματικού πεδίου ανάλυσης και οπτικοποίησης συναισθημάτων. Η εν λόγω ανασκόπηση περιλαμβάνει την παρουσίαση και επεξήγηση των πλέον χαρακτηριστικών τρόπων με τους οποίους μπορεί να αξιοποιηθεί η ανάλυση συναισθημάτων από επιχειρήσεις και οργανισμούς, ώστε να βελτιστοποιηθεί η εμπειρία που προσφέρεται στο κοινό, τους τελικούς πελάτες ή τους συναλλασσόμενους. Επιπρόσθετα, παρουσιάζεται μία λεπτομερής κατηγοριοποίηση των τεχνικών οπτικοποίησης συναισθημάτων, ομαδοποιημένων σε κατηγορίες ανάλογα με τα χαρακτηριστικά των χρησιμοποιούμενων δεδομένων, τις πηγές προέλευσης και ιδιότητες των δεδομένων, τις εργασίες που εκτελούνται και τις παραμέτρους και μεταβλητές οπτικοποίησης. Τέλος, έχει πραγματοποιηθεί αναλυτική καταγραφή εργαλείων και εφαρμογών (εμπορικών και ανοιχτού κώδικα), τα οποία δύναται να χρησιμοποιηθούν για την οπτικοποίηση αποτελεσμάτων ανάλυσης συναισθημάτων. Τα συγκεκριμένα εργαλεία χρησιμοποιούν πλήθος διαφορετικών τεχνικών ανάλυσης, μεταξύ των οποίων αλγόριθμους στατιστικής ανάλυσης και ομαδοποίησης (statistical analysis and clustering), τεχνητής νοημοσύνης (artificial intelligence) και μηχανικής και σε βάθος μάθησης (machine and deep learning), ενώ υφίσταται επίσης μεγάλο πλήθος εφαρμογών που έχουν αναπτυχθεί πάνω σε ανοιχτού κώδικα προγραμματιστικές πλατφόρμες και εργαλεία (π.χ. Python, R, κ.λπ.). Στο δεύτερο τμήμα της εργασίας παρουσιάζεται η μελέτη περίπτωσης (Case Study), στο πλαίσιο της οποίας έχει αναπτυχθεί διαδραστική διεπαφή, μέσω της οποίας παρουσιάζονται, αναλύονται και αξιολογούνται τα αποτελέσματα ανάλυσης συναισθημάτων. Για το σκοπό αυτό έχει χρησιμοποιηθεί ένα σύνολο δεδομένων (data set), το οποίο περιλαμβάνει κριτικές και αξιολογήσεις πελατών για προϊόντα και υπηρεσίες που παρέχονται από μια εταιρία πώλησης γυναικείων ρούχων. Για την ανάλυση συναισθημάτων και οπτικοποίηση των αποτελεσμάτων έχει χρησιμοποιηθεί η γλώσσα Python, περιλαμβανομένων και δημοφιλών βιβλιοθηκών γραφικής απεικόνισης.

Λέξεις κλειδιά: Ανάλυση συναισθημάτων, εξόρυξη γνώμης, οπτικοποίηση συναισθημάτων, Python.

Abstract

This study addresses the subject of sentiment analysis or opinion mining and in particular how the results of such an analysis can be presented so that they can be more easily understood and effectively utilized. In the first part of the Study, an extensive bibliographic review of the thematic field of sentiment analysis and visualization has been conducted. This review includes exploration and presentation of the most typical ways of which companies and organizations can take advantage to optimize customer and citizen experience. In addition, a detailed categorization of sentiment visualization techniques is presented, grouped into categories according to the characteristics of the data used, the sources and properties of the data, the tasks performed, and the visualization parameters and variables. Finally, a detailed presentation of tools and applications (commercial and open source), which can be used to visualize sentiment analysis results, has been conducted. These tools use a number of different analysis techniques, including statistical analysis and clustering algorithms, artificial intelligence, machine and deep learning algorithms, while there are also a large number of applications that have been developed on open source programming platforms and tools (e.g. Python, R, etc.). The second part of the thesis presents a case study in which an interactive graphical user interface has been developed, through which the results of sentiment analysis are presented, analyzed and evaluated. For this purpose, a data set has been used, which includes customer reviews for products and services provided by a women's clothing company. Python language has been utilized to implement sentiment analysis and visualize results, including popular graphics libraries.

Keywords: Sentiment analysis, opinion mining, sentiment analysis visualization, Python.

Πρόλογος – Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την κυρία Κολωνiάρη Γεωργία, επίκουρη καθηγήτρια ΠΑ.ΜΑΚ., για την ανάθεση αυτής της διπλωματικής εργασίας καθώς και για την υποστήριξη της κατά τη διάρκεια εκπόνησης της παρούσας διπλωματικής. Επίσης, θα ήθελα να ευχαριστήσω τον ξάδερφό μου Βλαχάβα Γεώργιο και την σύντροφό μου Τσιώπη Άννα για την στήριξη και την πολύτιμη βοήθειά τους κατά την διάρκεια των σπουδών μου.

Περιεχόμενα

Περίληψη	iv
Abstract.....	v
Πρόλογος – Ευχαριστίες.....	vi
Κατάλογος Σχημάτων και Εικόνων	ix
Κατάλογος Πινάκων.....	x
Συμβολισμοί	xi
1 Εισαγωγή	12
1.1 Πρόβλημα – Σημαντικότητα του θέματος.....	12
1.2 Σκοπός – Στόχοι	13
1.3 Συνεισφορά.....	15
1.4 Βασική Ορολογία	16
1.5 Διάρθρωση της μελέτης.....	17
2 Βιβλιογραφική Επισκόπηση – Θεωρητικό Υπόβαθρο	18
2.1 Εισαγωγή στην εξόρυξη γνώμης - ανάλυση συναισθημάτων	18
2.2 Τύποι ανάλυσης συναισθημάτων	19
2.3 Μέθοδοι και τεχνικές ανάλυσης συναισθημάτων	21
2.4 Εφαρμογές ανάλυσης συναισθημάτων	23
2.4.1 Μελέτη ανταγωνισμού επιχειρήσεων	23
2.4.2 Προσδιορισμός υποστηρικτών και επικριτών	24
2.4.3 Παρακολούθηση συναισθημάτων κοινού σε βάθος χρόνου.....	25
2.4.4 Ανάλυση αναφορών επωνυμίας επιχειρήσεων	25
2.4.5 Διεξαγωγή ερευνών αγοράς.....	26
2.5 Τεχνικές οπτικοποίησης ανάλυσης συναισθημάτων.....	27
2.5.1 Εισαγωγή στις τεχνικές οπτικοποίησης συναισθημάτων	27
2.5.2 Τεχνικές με βάση τα χαρακτηριστικά των δεδομένων	29
2.5.3 Τεχνικές με βάση την πηγή προέλευσης και ιδιότητες των δεδομένων	33
2.5.4 Τεχνικές με βάση τις εκτελούμενες εργασίες.....	35
2.5.5 Τεχνικές με βάση τα χαρακτηριστικά οπτικοποίησης	41
2.6 Εργαλεία και λογισμικό οπτικοποίησης ανάλυσης συναισθημάτων	46

3	Μελέτη Περίπτωσης: Ανάπτυξη Διαδραστικής Διεπαφής Παρουσίασης	
	Αποτελεσμάτων Ανάλυσης Συναισθημάτων.....	49
3.1	Εισαγωγή.....	49
3.2	Αντικείμενο υλοποίησης.....	49
3.3	Αναλυτικός σχεδιασμός διεπαφής – Δομή και βασικά Χαρακτηριστικά.....	51
3.4	Περιγραφή συνόλου δεδομένων	55
3.5	Εργαλεία ανάπτυξης εφαρμογής οπτικοποίησης στην Python.....	58
3.6	Παρουσίαση πρωτοτύπου (Prototype) υλοποίησης	62
3.7	Επεξήγηση – Ανάλυση Αποτελεσμάτων.....	63
4	Επίλογος.....	86
4.1	Συζήτηση αποτελεσμάτων	86
4.2	Σύνοψη και συμπεράσματα.....	88
4.3	Όρια και περιορισμοί της έρευνας – Μελλοντικές επεκτάσεις	91
	Βιβλιογραφία	93

Κατάλογος Σχημάτων και Εικόνων

Εικόνα 3-1: Λειτουργίες διάδρασης της διεπαφής.	54
Εικόνα 3-2: Βασικά χαρακτηριστικά της γλώσσας Python.	59
Εικόνα 3-3: Δημοφιλέστερες εφαρμογές της γλώσσας Python.	60
Εικόνα 3-4: Η πρώτη σελίδα της εφαρμογής οπτικοποίησης ανάλυσης συναισθημάτων.	70
Εικόνα 3-5: Κατανομή κριτικών με βάση το σκορ αξιολόγησης.	71
Εικόνα 3-6: Κατανομή κριτικών με βάση την πολικότητα της αξιολόγησης.	72
Εικόνα 3-7: Κατανομή κριτικών με βάση την ηλικία του πελάτη.	73
Εικόνα 3-8: Κατανομή κριτικών με βάση το μήκος κειμένου της αξιολόγησης.	74
Εικόνα 3-9: Κατανομή κριτικών με βάση τον αριθμό λέξεων της αξιολόγησης.	75
Εικόνα 3-10: Κατανομή κριτικών με βάση το δείκτη σύστασης.	76
Εικόνα 3-11: Οι δέκα (10) κορυφαίες λέξεις ανά μέγεθος ρούχου, δείκτη αξιολόγησης, σκορ αξιολόγησης και τύπο ρούχου, εξαιρουμένων των κοινών λέξεων (stop words).	77
Εικόνα 3-12: Τα δέκα (10) κορυφαία τρίλεξα ανά μέγεθος ρούχου, ηλικία, σκορ αξιολόγησης και τύπο ρούχου, εξαιρουμένων των κοινών λέξεων (stop words).	78
Εικόνα 3-13: Οι δέκα (10) κορυφαίες σημάσεις (tags) του κειμένου ανά μέγεθος ρούχου, ηλικία, σκορ αξιολόγησης και τύπο ρούχου, εξαιρουμένων των κοινών λέξεων (stop words).	79
Εικόνα 3-14: Κατανομή πολικότητας ανά τύπο ρούχου με αναπαράσταση boxplot.	80
Εικόνα 3-15: Κατανομή μήκους κριτικής ανά τύπο ρούχου με αναπαράσταση boxplot.	81
Εικόνα 3-16: Κατανομή πολικότητας ανά σκορ αξιολόγησης με αναπαράσταση boxplot.	82
Εικόνα 3-17: Κατανομή ηλικίας πελάτη ανά σκορ αξιολόγησης με αναπαράσταση boxplot.	83
Εικόνα 3-18: Λανθάνουσα σημασιολογική ανάλυση – Έξι (6) κορυφαία θέματα.	84
Εικόνα 3-19: Λανθάνουσα σημασιολογική ανάλυση (TSNE clustering) – Έξι (6) κορυφαία θέματα.	85

Κατάλογος Πινάκων

Πίνακας 2-1: Παραδείγματα λέξεων που αναδεικνύουν τον τύπο συναισθήματος.....	23
Πίνακας 2-2: Εργαλεία και λογισμικό οπτικοποίησης ανάλυσης συναισθημάτων.	48
Πίνακας 3-1: Πληροφορίες πηγαιού αρχείου συνόλου δεδομένων.	56
Πίνακας 3-2: Δείγμα πέντε (5) τυχαίων εγγραφών από το σύνολο δεδομένων.....	57
Πίνακας 3-3: Τα αρχεία της εφαρμογής και η λειτουργικότητά τους.	63

Συμβολισμοί

APIs	Application Programming Interfaces
FACS	Facial Action Coding System
NPL	Natural Language Processing
NPS	Net Promoter Score
POS	Part of Speech

1 Εισαγωγή

1.1 Πρόβλημα – Σημαντικότητα του θέματος

Οι επιχειρήσεις και οι οργανισμοί διαχρονικά επιδιώκουν να γνωρίζουν, να κατανοούν και να είναι σε θέση να αναλύσουν τα συναισθήματα των συναλλασσόμενων. Όσο περισσότερο περιορίζεται η φυσική παρουσία και η προσωπική επαφή τόσο σημαντικότερη καθίσταται η ανάγκη αναγνώρισης των συναισθημάτων, με τρόπους που αξιοποιούν την σύγχρονη τεχνολογία και τις δυνατότητές της. Παράλληλα, η ραγδαία ανάπτυξη των ψηφιακών τεχνολογιών και του διαδικτύου έχει οδηγήσει σε μια άνευ προηγουμένου αύξηση των δεδομένων κειμένου, δημιουργώντας έτσι νέες ευκαιρίες και προκλήσεις. Η έκρηξη του όγκου αλλά και ο βαθμός διακίνησης των δεδομένων, που σήμερα υπάρχουν διαθέσιμα από πολλαπλές πηγές και σε πολλαπλές μορφές, παρέχουν τεράστιες δυνατότητες κατανόησης των συναισθημάτων, αρκεί κανείς να είναι σε θέση να τα επεξεργαστεί αποτελεσματικά με μαζικό ή επιλεκτικό τρόπο, φιλτράροντας τον όποιο θόρυβο τα περιβάλλει. Επομένως, ένα από τα σύγχρονα ερευνητικά προβλήματα, που έχει αναδυθεί τα τελευταία χρόνια, είναι η προσπάθεια αυτοματοποιημένης ανάλυση συναισθημάτων (sentiment analysis), όρος που συναντάται και ως εξόρυξη γνώμης (opinion mining), και έχει ως στόχο την ανίχνευση και ανάλυση κειμένων και ταξινόμησή των γραφόμενων ανάλογα με τα συναισθήματα που εμπεριέχουν (Kucher et al., 2018; Liu, 2009; Pang & Lee, 2008).

Για περισσότερο από μία δεκαετία, η οπτικοποίηση των αναλύσεων συναισθημάτων και των στοιχείων, που αναδεικνύονται από την επεξεργασία των σχετικών κειμένων, συνιστά ένα σύγχρονο πεδίο ερευνητικής δραστηριότητας. Τα αποτελέσματα αναλύσεων συναισθημάτων δύναται πλέον να οπτικοποιηθούν με πολλούς και διαφορετικούς τρόπους, από παραδοσιακά γραφήματα πίτας ή ραβδογράμματα μέχρι εξειδικευμένα γραφήματα οπτικοποίησης, που περιλαμβάνουν νέους τρόπους αναπαράστασης, κατάλληλους για την ανάλυση πολύπλοκων και πολυδιάστατων συνόλων δεδομένων, που συνδυάζουν μεταξύ άλλων τις διαστάσεις χρόνου και χώρου και μπορούν την ίδια στιγμή να περιγράψουν αλληλοσυσχετίσεις (Kucher et al., 2018).

Διαφορετικά θεματικά πεδία εμπεριέχονται στην οπτικοποίηση συναισθημάτων ως ερευνητική περιοχή, όπως είναι η παρακολούθηση της κοινής γνώμης στα μέσα κοινωνικής

δικτύωσης, η ψηφιακή ανάλυση βιβλιογραφίας ανθρωπιστικών επιστημών, η έρευνα ανάλυσης συναισθημάτων, η γλωσσολογία και η επεξεργασία φυσικής γλώσσας (Natural Language Processing). Η οπτικοποίηση συναισθημάτων, κατά τις πρώτες φάσεις εξέλιξής της ως ερευνητικό πεδίο, συσχετίστηκε με την εξόρυξη δεδομένων (data mining) και την επεξεργασία φυσικής γλώσσας, ενώ οι τεχνικές οπτικοποίησης ήταν κατά κανόνα απλές. Οι τεχνικές, που σήμερα χρησιμοποιούνται, αντανακλούν την πρόοδο που έχει σημειωθεί τα τελευταία χρόνια στο ευρύτερο πεδίο της οπτικοποίησης, είναι πολύ πιο σύνθετες και διαχειρίζονται μεγάλο εύρος ετερογενών δεδομένων (Kucher et al., 2018).

1.2 Σκοπός – Στόχοι

Η παρούσα εργασία ερευνά το θέμα της ανάλυσης συναισθημάτων και ειδικότερα τους τρόπους παρουσίασης των αποτελεσμάτων μιας τέτοιας ανάλυσης, προκειμένου να καταστούν ευκολότερα κατανοητά και να αξιοποιηθούν με αποτελεσματικότερο τρόπο. Οι βασικοί στόχοι της παρούσας έρευνας είναι οι ακόλουθοι:

- 1) Ευρεία βιβλιογραφική ανασκόπηση του θεματικού πεδίου ανάλυσης και οπτικοποίησης συναισθημάτων, με τους ακόλουθους επιμέρους στόχους:
 - i. Παρουσίαση και επεξήγηση των πλέον χαρακτηριστικών τρόπων, με τους οποίους μπορεί να αξιοποιηθεί η ανάλυση συναισθημάτων από επιχειρήσεις και οργανισμούς, ώστε να βελτιστοποιηθεί η εμπειρία που προσφέρεται στο κοινό, τους τελικούς πελάτες ή τους συναλλασσόμενους. Σημαντικές περιοχές, στις οποίες βρίσκει εφαρμογή η ανάλυση συναισθημάτων, είναι η μελέτη του ανταγωνισμού μιας επιχείρησης και ο προσδιορισμός υποστηρικτών και επικριτών επί ενός θέματος, ή μιας άποψης.
 - ii. Παρουσίαση και λεπτομερής ανασκόπηση και κατηγοριοποίηση των τεχνικών οπτικοποίησης συναισθημάτων, ομαδοποιημένων σε κατηγορίες ανάλογα με τα χαρακτηριστικά των χρησιμοποιούμενων δεδομένων, τις πηγές προέλευσης και ιδιότητες των δεδομένων, τις εργασίες που εκτελούνται και τις παραμέτρους και μεταβλητές οπτικοποίησης.
 - iii. Καταγραφή εργαλείων και εφαρμογών (εμπορικών και ανοιχτού κώδικα – open source), τα οποία δύναται να χρησιμοποιηθούν για την οπτικοποίηση αποτελεσμάτων

ανάλυσης συναισθημάτων. Τα συγκεκριμένα εργαλεία χρησιμοποιούν πλήθος διαφορετικών τεχνικών ανάλυσης, μεταξύ των οποίων αλγόριθμους στατιστικής ανάλυσης και ομαδοποίησης (statistical analysis and clustering), τεχνητής νοημοσύνης (artificial intelligence) και μηχανικής και σε βάθος μάθησης (machine and deep learning), ενώ υφίσταται επίσης μεγάλο πλήθος εφαρμογών, που έχουν αναπτυχθεί πάνω σε ανοιχτού κώδικα προγραμματιστικές πλατφόρμες και εργαλεία (π.χ. Python, R, κ.λπ.).

- 2) Μελέτη περίπτωσης (Case Study), στην οποία έχει αναπτυχθεί διαδραστική διεπαφή μέσω της οποίας παρουσιάζονται, αναλύονται και αξιολογούνται τα αποτελέσματα ανάλυσης συναισθημάτων. Για το σκοπό αυτό έχει χρησιμοποιηθεί ένα σύνολο δεδομένων (data set), το οποίο περιλαμβάνει κριτικές και αξιολογήσεις πελατών για προϊόντα και υπηρεσίες που παρέχονται από μια εταιρία πώλησης γυναικείων ρούχων. Για την ανάλυση συναισθημάτων και οπτικοποίηση των αποτελεσμάτων έχει χρησιμοποιηθεί η γλώσσα Python, περιλαμβανομένων και δημοφιλών βιβλιοθηκών γραφικής απεικόνισης όπως είναι η Plotly και η Dash.

Η διεπαφή είναι προσπελάσιμη και απεικονίζει τα γραφικά μέσω περιηγητή ιστοσελίδων (web browser). Για το κείμενο που αναλύεται, απεικονίζονται πολλαπλά γραφικά και πληροφορίες, τα οποία συνδυάζονται και ομαδοποιούνται με λογικό τρόπο ανά σελίδα. Η εφαρμογή είναι πολλαπλών σελίδων (multi page application), οι οποίες έχουν μια λογική αλληλουχία. Οι λογικές ενότητες/ σελίδες της web-εφαρμογής, που ομαδοποιούν τις πληροφορίες και τα γραφικά τα οποία απεικονίζονται μαζί, είναι οι ακόλουθες:

- Ενότητα Α – Data Sample Basics: Εισαγωγική σελίδα, η οποία περιλαμβάνει βασικές πληροφορίες για το σύνολο δεδομένων (data set).
- Ενότητα Β - Visualizations: Ομάδα ραβδογραμμάτων (bar charts), που οπτικοποιούν τη στατιστική ανάλυση των κριτικών και την κατανομή τους βάσει διαφορετικών κριτηρίων.
- Ενότητα Γ – Box Plots: Γραφήματα τύπου box plot, που απεικονίζουν με ιδιαίτερα παραστατικό και διαδραστικό τρόπο τα συναισθήματα βάσει διαφορετικών κριτηρίων.

- Ενότητα Δ – Latent Analysis: Γραφική απεικόνιση λανθάνουσας ανάλυσης (latent analysis), που παρουσιάζει τα έξι (6) δημοφιλέστερα θέματα συζήτησης (topics) μέσα από δύο διαφορετικούς τύπους γραφημάτων.

1.3 Συνεισφορά

Η παρούσα εργασία μπορεί να είναι καταρχήν επωφελής για τους ερευνητές οπτικοποίησης συναισθημάτων, καθώς συγκεντρώνει και παρουσιάζει μια ευρεία βιβλιογραφική ανάλυση της συγκεκριμένης περιοχής, χρησιμοποιώντας ως βάση δημοσιεύσεις, που είναι διάσπαρτες σε μεγάλο αριθμό επιστημονικών υποπεριοχών. Επιπρόσθετα, μπορεί να είναι χρήσιμη και για ερευνητές άλλων επιστημονικών πεδίων αλλά και για επαγγελματίες, που ενδιαφέρονται για μεθόδους οπτικής ανάλυσης δεδομένων συναισθημάτων, αφού πρόκειται για μια αγορά που αποτιμάται περί τα 22.65 δισεκατομμύρια δολάρια.

Οι περιοχές συνεισφοράς της παρούσας εργασίας συνοψίζονται στα ακόλουθα:

- Ευρεία βιβλιογραφική ανασκόπηση του θεματικού πεδίου ανάλυσης και οπτικοποίησης συναισθημάτων και της έρευνας που έχει πραγματοποιηθεί στη συγκεκριμένη περιοχή.
- Επισκόπηση ενός ευρύτατου συνόλου τεχνικών οπτικοποίησης συναισθημάτων, που έχουν μελετηθεί και εφαρμοστεί σε πολλούς διαφορετικούς ερευνητικούς κλάδους την τελευταία εικοσαετία.
- Λεπτομερής κατηγοριοποίηση και ανάλυση των χρησιμοποιούμενων τεχνικών οπτικοποίησης συναισθημάτων. Διερεύνηση υφιστάμενων τάσεων και αλληλοσυσχετίσεων μεταξύ των επιμέρους κατηγοριών οπτικοποίησης συναισθημάτων.
- Πρακτικές υλοποίησης οπτικοποίησης ανάλυσης συναισθημάτων, μέσα από την υλοποίηση σχετικής μελέτης περίπτωσης, ώστε να γίνει εύκολα αντιληπτό πώς μπορούν να αξιοποιηθούν εργαλεία ανοιχτού κώδικα στη συγκεκριμένη θεματική περιοχή.
- Προτάσεις για μελλοντικές ερευνητικές κατευθύνσεις στην περιοχή της οπτικοποίησης συναισθημάτων.

1.4 Βασική Ορολογία

Επιχειρώντας την απόδοση ενός τεχνικού ορισμού της ανάλυσης συναισθημάτων ή εξόρυξης γνώμης, δύναται να σημειωθεί ότι πρόκειται για μια ανάλυση δεδομένων κειμένου, η οποία μπορεί να καταδείξει την «πολικότητα» των συναισθημάτων που εμπεριέχονται, δηλαδή θετικά, ουδέτερα ή αρνητικά συναισθήματα, που μπορούν επιπρόσθετα να ποσοτικοποιηθούν με κάποιο score. Η ανάλυση συναισθημάτων προκύπτει από την αξιολόγηση των απόψεων, που εμπεριέχουν τα δεδομένα κειμένου, είναι επομένως γνωστή και ως εξόρυξη γνώμης. Ένας εναλλακτικός ορισμός για την ανάλυση συναισθημάτων θα ήταν ότι συνιστά μια διαδικασία εντοπισμού και υπολογιστικής ανάλυσης απόψεων και κρίσεων, οι οποίες διατυπώνονται σε τμήματα κειμένου, βάσει της οποίας συνάγεται αν τα υπό εξέταση τμήματα κειμένου εκφράζουν θετικές, ουδέτερες, ή αρνητικές απόψεις (Pang & Lee, 2008).

Ως όρος η «*ανάλυση συναισθημάτων*» συνιστά συνήθως τη διαδικασία κατηγοριοποίησης τμημάτων κειμένου, που εκτείνονται από μεμονωμένες λέξεις και φράσεις ή προτάσεις έως πλήρη έγγραφα, σε ένα μικρό αριθμό κατηγοριών που αντιπροσωπεύουν διαφορετικά είδη συναισθημάτων. Ο όρος «*συναίσθημα*», στη συγκεκριμένη επιστημονική περιοχή, λαμβάνεται συνήθως ως συνώνυμο όρων όπως συγκίνηση, επίδραση, στάση κ.λπ. Στην απλούστερη διατύπωσή της, η ανάλυση συναισθημάτων λογίζεται δυαδικό πρόβλημα, κατά την επίλυση του οποίου εντοπίζονται και διαχωρίζονται συναισθηματικά τα θετικά από το αρνητικά συναισθήματα, διαδικασία γνωστή και ως ανίχνευση πολικότητας (Turney, 2002). Σε ερευνητικές προσπάθειες που ακολούθησαν κυριάρχησε η προσέγγιση κατά την οποία στην ανάλυση συναισθημάτων πραγματοποιείται κατηγοριοποίησή τους σε τρεις κατηγορίες, αρνητικά, ουδέτερα και θετικά, ενώ υπάρχει και βιβλιογραφία, που χρησιμοποιεί κλιμακωτή διαβάθμιση σε περισσότερες από τρεις κατηγορίες, για παράδειγμα ασθενώς έως έντονα αρνητική, ουδέτερη, ασθενώς ως έντονα θετική (Socher et al., 2013; Socher et al., 2011; He, 2012; Amigó et al., 2013).

Με τον όρο «*οπτικοποίηση συναισθημάτων*» (Sentiment Visualization) γίνεται αναφορά στους τρόπους και τις τεχνικές, μέσα από τις οποίες δύναται να απεικονιστούν και να αναλυθούν οι πληροφορίες ανάλυσης συναισθημάτων, που προέκυψαν από τη συλλογή και επεξεργασία δεδομένων κειμένου. Η οπτικοποίηση συναισθημάτων συνιστά υποσύνολο της ευρύτερης ερευνητικής περιοχής, που είναι γνωστή ως οπτικοποίηση κειμένου (text visualization) [Kucher & Kerren, 2015].

1.5 Διάρθρωση της μελέτης

Το παρόν κεφάλαιο περιλαμβάνει καταρχάς τη διατύπωση του προβλήματος που πραγματεύεται η παρούσα μελέτη (Ενότητα 1.1). Οι βασικοί στόχοι και η συνεισφορά της εργασίας περιέχονται αντίστοιχα στις Ενότητες 1.2 και 1.3, ενώ στην Ενότητα 1.4 παρατίθενται οι βασικοί ορισμοί εννοιών. Το Κεφάλαιο 2 παρουσιάζει μια ευρεία βιβλιογραφική επισκόπηση του αντικειμένου μαζί με το θεωρητικό υπόβαθρο που απαιτείται για την κατανόηση της συγκεκριμένης ερευνητικής περιοχής. Ειδικότερα, στην Ενότητα 2.1 επιχειρείται μια εισαγωγή στην εξόρυξη γνώμης, στην Ενότητα 2.2 παρουσιάζονται οι τύποι ανάλυσης συναισθημάτων και στην Ενότητα 2.3 παρατίθενται και περιγράφονται οι βασικές μέθοδοι και τεχνικές ανάλυσης συναισθημάτων. Στην Ενότητα 2.4 επιχειρείται μια αναλυτική καταγραφή των εφαρμογών ανάλυσης συναισθημάτων, ενώ η Ενότητα 2.5 εισάγει τις τεχνικές οπτικοποίησης που χρησιμοποιούνται στην ανάλυση συναισθημάτων. Το κεφάλαιο κλείνει με την Ενότητα 2.6, η οποία εξετάζει τα εργαλεία και το λογισμικό οπτικοποίησης που χρησιμοποιούνται για την ανάλυση συναισθημάτων.

Το Κεφάλαιο 3 παρουσιάζει και αναλύει τον τρόπο με τον οποίο υλοποιήθηκε η μελέτη περίπτωσης της παρούσας εργασίας και συγκεκριμένα την ανάπτυξη μιας διαδραστικής διεπαφής για την παρουσίαση αποτελεσμάτων ανάλυσης συναισθημάτων. Ειδικότερα, μετά την εισαγωγή της Ενότητας 3.1, παρατίθεται μία αναλυτική περιγραφή του αντικειμένου υλοποίησης (Ενότητα 3.2) και εν συνεχεία ο αναλυτικός σχεδιασμός της διεπαφής μαζί με τη δομή και τα βασικά χαρακτηριστικά της (Ενότητα 3.3) και η δομή και περιγραφή του συνόλου δεδομένων που χρησιμοποιείται (Ενότητα 3.4). Η Ενότητα 3.5 απαριθμεί και επεξηγεί το λογισμικό και τα εργαλεία ανάπτυξης εφαρμογής οπτικοποίησης, πάνω στα οποία εδράζεται η μελέτη περίπτωσης και ειδικότερα τη γλώσσα Python και τις βιβλιοθήκες γραφικών που χρησιμοποιούνται. Στην Ενότητα 3.6 παρουσιάζεται το υλοποιηθέν πρωτότυπο, ενώ στην Ενότητα 3.7 επεξηγούνται και αναλύονται τα αποτελέσματα που παράγει η υλοποιηθείσα εφαρμογή.

Το Κεφάλαιο 4 συνιστά την ολοκλήρωση της εργασίας και ειδικότερα μία σύνοψη και τα συμπεράσματα που προέκυψαν στην Ενότητα 4.1, τα όρια και τους περιορισμούς της πραγματοποιηθείσας έρευνας καθώς επίσης και τις προτεινόμενες μελλοντικές επεκτάσεις στην Ενότητα 4.2. Τέλος, ακολουθεί η βιβλιογραφία.

2 Βιβλιογραφική Επισκόπηση – Θεωρητικό Υπόβαθρο

2.1 Εισαγωγή στην εξόρυξη γνώμης - ανάλυση συναισθημάτων

Παρά το γεγονός ότι στη σημερινή εποχή πολλές φορές οι συναλλαγές του κοινού ή των πελατών γίνονται χωρίς να απαιτείται φυσική παρουσία, η ανάγκη για τις επιχειρήσεις και τους οργανισμούς να γνωρίζουν, να κατανοούν και να μπορούν να αναλύσουν τα συναισθήματα των συναλλασσόμενων παραμένει σταθερά υψηλή διαχρονικά. Μάλιστα, όσο περισσότερο περιορίζεται η φυσική παρουσία και η προσωπική επαφή τόσο σημαντικότερη καθίσταται η ανάγκη αναγνώρισης των συναισθημάτων με τρόπους που αξιοποιούν την σύγχρονη τεχνολογία και τις δυνατότητές της. Από την άλλη πλευρά, η ραγδαία ανάπτυξη των ψηφιακών τεχνολογιών και του διαδικτύου οδήγησε σε μια άνευ προηγουμένου αύξηση των δεδομένων κειμένου, δημιουργώντας νέες ευκαιρίες και προκλήσεις. Οι ερευνητές γλωσσολόγοι έχουν πλέον πρόσβαση στα δεδομένα, που μπορεί να διαφέρουν δραστικά ανάλογα με το αντικείμενο ή το θεματικό πεδίο όσον αφορά στο περιεχόμενο και την ευρύτητα (Kucher et al., 2018).

Δεδομένης της έκρηξης του όγκου αλλά και του βαθμού διακίνησης των δεδομένων, σήμερα υπάρχουν διαθέσιμα από πολλαπλές πηγές και σε πολλαπλές μορφές δεδομένα, τα οποία παρέχουν τεράστιες δυνατότητες κατανόησης των συναισθημάτων, αρκεί κανείς να επιτύχει την αποτελεσματική επεξεργασία τους με μαζικό ή επιλεκτικό τρόπο, φιλτράροντας τον όποιο θόρυβο τα περιβάλλει. Ως εκ τούτου, ένα από τα σύγχρονα ερευνητικά προβλήματα που έχει αναδυθεί τα τελευταία χρόνια είναι η προσπάθεια αυτοματοποιημένης ανάλυση συναισθημάτων (sentiment analysis), όρος που συναντάται και ως εξόρυξη γνώμης (opinion mining), και έχει ως στόχο την ανίχνευση και ανάλυση κειμένων και ταξινόμησή των γραφόμενων ανάλογα με τα συναισθήματα που εμπεριέχουν (Liu, 2009; Pang & Lee, 2008).

Η παρούσα εργασία διερευνά το ζήτημα της ανάλυσης συναισθημάτων και ειδικότερα το πώς μπορούν να παρουσιαστούν τα αποτελέσματα μια τέτοιας ανάλυσης, ώστε να μπορούν να γίνουν ευκολότερα κατανοητά και να αξιοποιηθούν με αποτελεσματικότερο τρόπο. Εάν επιχειρήσει κάποιος να δώσει έναν τεχνικό ορισμό στην ανάλυση συναισθημάτων ή εξόρυξη γνώμης, θα μπορούσε να πει ότι πρόκειται για μια ανάλυση δεδομένων κειμένου, που μπορεί να καταδείξει την «πολικότητα» των συναισθημάτων που εμπεριέχονται, δηλαδή θετικά, ουδέτερα ή

αρνητικά συναισθήματα, που μπορούν επιπρόσθετα να ποσοτικοποιηθούν με κάποιο score. Δεδομένου ότι η ανάλυση συναισθημάτων προκύπτει από την αξιολόγηση των απόψεων που εμπεριέχουν τα δεδομένα κειμένου, είναι επίσης γνωστή και ως εξόρυξη γνώμης. Εναλλακτικά, θα μπορούσε κανείς να ορίσει την ανάλυση συναισθημάτων ως μια διαδικασία εντοπισμού και υπολογιστικής ανάλυσης απόψεων και κρίσεων που διατυπώνονται σε τμήματα κειμένου, βάσει της οποίας προκύπτει αν τα υπό εξέταση τμήματα κειμένου εκφράζουν θετικές, ουδέτερες, ή αρνητικές απόψεις (Pang & Lee, 2008).

Αν και η ανάλυση συναισθημάτων έχει γίνει ιδιαίτερα δημοφιλής τα τελευταία χρόνια, τα πρώτα βήματα θεμελίωσης της εν λόγω επιστημονικής περιοχής ξεκίνησαν από τις αρχές της δεκαετίας του 2000. Κατά τη διάρκεια αυτής της περιόδου, υπήρξαν πολλές εργασίες και προσπάθειες, οι οποίες επικεντρώθηκαν σε σημαντικό βαθμό στον τρόπο συλλογής και ανάλυσης δεδομένων. Ένας από τους σπουδαιότερους λόγους, για τον οποίο η ανάλυση συναισθημάτων είναι σημαντική, είναι ότι δίνει τη δυνατότητα σε όποιον έχει ενδιαφέρον να το κάνει, να κατανοήσει και να αναλύσει τη γενική γνώμη του κοινού για ένα ζήτημα, ένα προϊόν, μια υπηρεσία, κ.λπ. (Pang & Lee, 2008; Norambuena et al., 2019).

Η μεγαλύτερη πρόκληση που παρουσιάζει η ανάλυση συναισθημάτων πηγάζει από τις ιδιαιτερότητες της εκάστοτε φυσικής γλώσσας, οι οποίες δύνανται να διαφοροποιήσουν το νόημα και την ερμηνεία κατά περίπτωση και ανάλογα με τα συμφραζόμενα, ειδικότερα δε σε περιπτώσεις που υφίσταται ειρωνεία ή σαρκασμός (Liu, 2009).

2.2 Τύποι ανάλυσης συναισθημάτων

Ο τρόπος με τον οποίο χρησιμοποιείται κάθε φορά η φυσική γλώσσα διαφοροποιείται ανάλογα με τους συμμετέχοντες, την εκάστοτε οπτική τους και τον τρόπο με τον οποίο αλληλεπιδρούν. Ο όρος ανάλυση συναισθημάτων ορίζεται συνήθως ως η διαδικασία κατηγοριοποίησης τμημάτων κειμένου, που εκτείνονται από μεμονωμένες λέξεις και φράσεις ή προτάσεις έως πλήρη έγγραφα, σε ένα μικρό αριθμό κατηγοριών που αντιπροσωπεύουν διαφορετικά είδη συναισθημάτων. Ο όρος συναίσθημα στη συγκεκριμένη επιστημονική περιοχή αποτελεί συνήθως συνώνυμο όρων όπως συγκίνηση, επίδραση, στάση κ.λπ. Στην απλούστερη διατύπωσή της, η ανάλυση συναισθημάτων θεωρείται δυαδικό πρόβλημα, κατά την επίλυση του οποίου εντοπίζονται και διαχωρίζονται συναισθηματικά τα θετικά από το αρνητικά

συναισθήματα, διαδικασία γνωστή και ως ανίχνευση πολικότητας (Turney, 2002). Σε μεταγενέστερες ερευνητικές προσπάθειες, κυριάρχησε η προσέγγιση κατά την οποία στην ανάλυση συναισθημάτων πραγματοποιείται κατηγοριοποίησή τους σε τρεις κατηγορίες, αρνητικά, ουδέτερα και θετικά, ενώ υπάρχει και βιβλιογραφία που χρησιμοποιεί κλιμακωτή διαβάθμιση σε περισσότερες από τρεις κατηγορίες, για παράδειγμα ασθενώς έως έντονα αρνητική, ουδέτερη, ασθενώς ως έντονα θετική (Socher et al., 2013; Socher et al., 2011; He, 2012; Amigó et al., 2013).

Υπάρχει ευρεία έρευνα σε εξέλιξη στη συγκεκριμένη περιοχή και επομένως υπάρχουν στη βιβλιογραφία διαφορετικές κατηγοριοποιήσεις των τύπων ανάλυσης συναισθημάτων. Θα μπορούσε κανείς να συνοψίσει τους τύπους ανάλυσης συναισθημάτων στις εξής τέσσερις κατηγορίες (Hogenboom et al., 2014; Norambuena et al., 2019; Taboada et al., 2011; Kucher et al., 2018):

- Λεπτομερής Ανάλυση Συναισθήματος (Fine-grained Sentiment Analysis): Πρόκειται για ανάλυση με την οποία μπορούν να γίνουν κατανοητά το ύφος και η πολικότητα των σχολίων επί κάποιου θέματος. Ο συγκεκριμένος τύπος ανάλυσης είναι εν γένει απαιτητικός και υψηλού κόστους σε σύγκριση με άλλα είδη, διότι διενεργείται σε μεγάλο επίπεδο λεπτομέρειας.
- Ανάλυση Ανίχνευσης Συναισθήματος (Emotion Detection Sentiment Analysis): Είναι προηγμένος τύπος ανάλυσης και αναγνώρισης συναισθήματος ο οποίος αξιοποιεί τεχνικές όπως χρήση λεξικών και μηχανική μάθηση (machine learning). Ένα από τα πιο γνωστά μοντέλα αυτού του τύπου είναι το μοντέλο «Big Six» του Ekman, το οποίο διαβαθμίζει τα αισθήματα σε θυμό, φόβο, ευτυχία, έκπληξη, αηδία και θλίψη (Ekman, 1992). Επίσης, στην κατηγορία αυτή κατατάσσονται τα διαστατικά (dimensional) μοντέλα, τα οποία κατηγοριοποιούν τα συναισθήματα σε περιοχές όπως ευχαρίστηση, διέγερση, κυριαρχία, κ.λπ. (Russell, & Mehrabian, 1977). Ειδικότερα, τα λεξικά περιλαμβάνουν λίστες λέξεων με θετική ή αρνητική ερμηνεία, ώστε να καθίσταται εφικτή η κατανόηση του συναισθήματος, ενώ για την συνολική αξιολόγηση ενός κειμένου χρησιμοποιούνται αλγόριθμοι τεχνητής νοημοσύνης και μηχανικής μάθησης. Παραδείγματα τέτοιων λεξικών είναι τα MPQA Subjectivity Lexicon (Wilson et al., 2005), WordNet-Affect (Srapparava & Valitutti, 2004), SentiWordNet (Baccianella et al., 2010), LIWC (Tausczik & Pennebaker, 2010), SenticNet (Cambria et al., 2014), κ.λπ.

- Μονοδιάστατη Ανάλυση Συναισθήματος (Aspect-based Sentiment Analysis): Πρόκειται για τύπο ανάλυσης συναισθήματος που επικεντρώνεται σε μία πτυχή μιας υπηρεσίας ή ενός προϊόντος. Για παράδειγμα, για μια εταιρεία που πουλά έπιπλα θα μπορούσε να επιχειρηθεί ανάλυση συναισθήματος μόνο για το σχήμα ή το χρώμα τους (Brody & Elhadad, 2010; Jo & Oh, 2011).
- Ανάλυση Προθέσεων (Intent Analysis): Ο συγκεκριμένος τύπος ανάλυσης συναισθήματος εμβαθύνει την ανάλυση σε τέτοιο βαθμό ώστε να μπορεί να δώσει και προβλέψεις της πρόθεσης ενός ατόμου με βάση τα χαρακτηριστικά του συναισθήματός του. Για παράδειγμα, χρησιμοποιώντας αυτή την τεχνική, μία εταιρεία θα μπορούσε να προβλέψει αν ένας πελάτης προτίθεται να αγοράσει ή να χρησιμοποιήσει ένα προϊόν ή μια υπηρεσία.

2.3 Μέθοδοι και τεχνικές ανάλυσης συναισθημάτων

Η ανάλυση συναισθημάτων επιτυγχάνεται τεχνικά όχι μόνο με την εξέταση κάθε δεδομένου/ λέξης μεμονωμένα, αλλά επίσης και συνδυαστικά ώστε να συνδεθούν και να ερμηνευτούν με ολιστικό τρόπο όλες οι έννοιες που έχουν καταγραφεί και η ερμηνεία τους. Τα δεδομένα κειμένου συνήθως προέρχονται από διάφορες πηγές όπως μέσα κοινωνικής δικτύωσης, μηνύματα (email, sms, chat, κ.λπ.), ιστότοπους, μεταδεδομένα (εικόνων, βίντεο, κ.λπ.) και άλλα (Pang & Lee, 2008; Mohammad, 2016; Kucher et al., 2018).

Για την υλοποίηση των προαναφερθέντων τύπων ανάλυσης συναισθήματος χρησιμοποιούνται διαφορές τεχνικές. Στην απλή περίπτωση οι τεχνικές αυτές στηρίζονται σε κανόνες (rule-based), ενώ στους πλέον προηγμένους τύπους ανάλυσης, όπως ήδη αναφέρθηκε, χρησιμοποιούνται αυτοματοποιημένες τεχνικές που στηρίζονται σε αλγόριθμους τεχνικής νοημοσύνης (artificial intelligence) και μηχανικής μάθησης (machine learning). Ακολουθώς παρατίθενται οι βασικές τεχνικές ανάλυσης συναισθημάτων (Pang & Lee, 2008; Mohammad, 2016):

- Τεχνικές Ανάλυσης Συναισθήματος που Στηρίζονται σε Κανόνες (Rule-based): Καθορίζονται από τον προγραμματιστή με ντετερμινιστικό τρόπο. Η ανάλυση στηρίζεται σε απλούς ή σύνθετους κανόνες, η εκτέλεση των οποίων καθορίζει και το αποτέλεσμα. Παράδειγμα τέτοιου κανόνα θα μπορούσε να είναι ο υπολογισμός ενός score με βάση τον

συνολικό αριθμό των θετικών και αρνητικών λέξεων που εμφανίζονται σε ένα κείμενο. Το ύψος του score που προκύπτει με τον τρόπο αυτό καθορίζει και το αν τελικά το υπό εξέταση κείμενο θα χαρακτηριστεί «θετικό», «ουδέτερο» ή «αρνητικό».

Οι rule-based τεχνικές δεν ενδείκνυνται για περιπτώσεις κατά τις οποίες τα δεδομένα μεταβάλλονται δυναμικά ή έχουν μεγάλο βαθμό πολυπλοκότητας.

- Τεχνικές Μηχανικής Μάθησης: Εφαρμόζουν αλγόριθμους που δεν απαιτούν χειροκίνητο (manual) χειρισμό. Κατά την εφαρμογή αυτών των τεχνικών οι αλγόριθμοι που χρησιμοποιούνται εκπαιδεύουν το σύστημα να αναγνωρίζει με τρόπο δυναμικό τα δεδομένα και να τα διαβαθμίζει ως θετικά, ουδέτερα ή αρνητικά. Επομένως, οι εν λόγω τεχνικές ενδείκνυνται για αναλύσεις μεγαλύτερης πολυπλοκότητας. Χαρακτηριστικά παραδείγματα αλγορίθμων μηχανικής μάθησης που χρησιμοποιούνται για ανάλυση συναισθημάτων είναι η γραμμική παλινδρόμηση (linear regression), ο αλγόριθμος Naive Bayes, οι τεχνικές διανυσματικής μηχανικής (vector machines) και αλγόριθμοι βαθιάς μάθησης (deep learning).
- Υβριδικές Τεχνικές: Όταν οι rule-based τεχνικές συνδυάζονται με τεχνικές μηχανικής μάθησης, προκύπτουν τεχνικές που συχνά αποκαλούνται υβριδικές (hybrid). Το πλεονέκτημα αυτών των τεχνικών είναι ότι ισορροπούν τον απαιτούμενο αυτοματισμό με την πολυπλοκότητα υλοποίησης και ενδείκνυνται στην περίπτωση που αφενός οι rule-based τεχνικές δε δουλεύουν καλά και αφετέρου οι τεχνικές μηχανικής μάθησης είναι ιδιαίτερα πολύπλοκες χωρίς να φέρνουν ανάλογη της πολυπλοκότητας αυτής προστιθέμενη αξία.

2.4 Εφαρμογές ανάλυσης συναισθημάτων

Στην ενότητα αυτή παρουσιάζονται και επεξηγούνται οι πλέον χαρακτηριστικοί τρόποι, με τους οποίους μπορεί να αξιοποιηθεί η ανάλυση συναισθημάτων από επιχειρήσεις και οργανισμούς, ώστε να βελτιστοποιηθεί η εμπειρία που προσφέρεται στο κοινό, τους τελικούς πελάτες ή τους συναλλασσόμενους. Οι σημαντικότερες περιοχές, στις οποίες βρίσκει εφαρμογή η ανάλυση συναισθημάτων, είναι η μελέτη του ανταγωνισμού μιας επιχείρησης και ο προσδιορισμός υποστηρικτών και επικριτών επί ενός θέματος, ή μιας άποψης, τα οποία και παρουσιάζονται στη συνέχεια της παρούσας ενότητας.

2.4.1 Μελέτη ανταγωνισμού επιχειρήσεων

Έχει πλέον γίνει πάγια πρακτική στο σύγχρονο ανταγωνιστικό επιχειρείν οι επιχειρήσεις και οργανισμοί να παρακολουθούν στενά τους ανταγωνιστές τους και να προσπαθούν να εκμεταλλευτούν τα κενά, τις αδυναμίες και τις ελλείψεις τους, ώστε να ενισχύσουν τη δική τους επιχειρηματική δραστηριότητα, τη διεύρυνση του πελατολογίου της και την αύξηση των εσόδων και της κερδοφορίας τους. Ένας από τους δημοφιλέστερους τρόπους για να επιτύχουν κάτι τέτοιο είναι η παρακολούθηση των σχολίων, απόψεων, τοποθετήσεων και αξιολογήσεων των πελατών ανταγωνιστών όπως αυτά καταγράφονται και δημοσιοποιούνται σε δημόσια προσβάσιμες πηγές πληροφόρησης, π.χ. Facebook, Twitter, Google, κ.λπ. (Kucher et al., 2018).

Για παράδειγμα, αν κάποιος χρησιμοποιήσει ως δεδομένα τις αξιολογήσεις των πελατών ενός ανταγωνιστικού ηλεκτρονικού καταστήματος για ένα συγκεκριμένο προϊόν και επιχειρήσει βάσει αυτών την ανάλυση συναισθημάτων των πελατών που έχουν πραγματοποιήσει τις εν λόγω αξιολογήσεις, μπορούν να γίνουν κατανοητά τα βασικά σημεία που αρέσουν ή απογοητεύουν τους πελάτες του ανταγωνισμού. Η ανάλυση συναισθημάτων στην προκειμένη περίπτωση μπορεί να αναδείξει θετικά, ουδέτερα ή αρνητικά συναισθήματα βάσει των λέξεων που κάθε φορά χρησιμοποιούν οι πελάτες. Παραδείγματα λέξεων που αναδεικνύουν τον τύπο συναισθήματος στο συγκεκριμένο παράδειγμα θα μπορούσαν να είναι αυτά που εμφανίζονται στον παρακάτω πίνακα.

<i>Εύχρηστο</i>	<i>Πολύπλοκο</i>	<i>Μέτριο</i>
<i>Κακή υποστήριξη</i>	<i>Ακριβό</i>	<i>Κακή επιλογή</i>
<i>Ωραία εμφάνιση</i>	<i>Υψηλό κόστος</i>	<i>Καλή ποιότητα</i>

Πίνακας 2-1: Παραδείγματα λέξεων που αναδεικνύουν τον τύπο συναισθήματος.

2.4.2 Προσδιορισμός υποστηρικτών και επικριτών

Είναι κοινή πρακτική επιχειρήσεων και οργανισμών, που επιθυμούν να αξιολογήσουν και να ενισχύσουν την πιστότητα των πελατών τους, να συλλέγουν και να αναλύουν τις απόψεις και τα σχόλιά τους σχετικά με προϊόντα ή υπηρεσίες που παρέχουν. Συνήθεις τρόποι υποβολής απόψεων και σχολίων αποτελούν τα ερωτηματολόγια, οι έρευνες αξιολόγησης της σχέσης που έχει αναπτυχθεί μεταξύ μίας εταιρίας ή ενός brand και των πελατών της (Net Promoter Score, NPS), οι δημοσκοπήσεις που πραγματοποιούνται στα μέσα κοινωνικής δικτύωσης, οι αξιολογήσεις, οι τηλεφωνικές συνομιλίες, τα μηνύματα (email, sms, chat), κ.λπ. Μέσα από τέτοιου τύπου έρευνες και αναλύσεις, το ζητούμενο είναι να προσδιορίζονται τόσο οι υποστηρικτές (promoters) όσο και οι επικριτές (detractors) της επιχείρησης, ώστε μπορούν να ληφθούν αποφάσεις για μελλοντικές επιχειρηματικές κινήσεις. Ως υποστηρικτές χαρακτηρίζονται οι πελάτες εκείνοι που παραμένουν πιστοί στην επωνυμία μιας επιχείρησης και μάλιστα είναι αυτοί που συστήνουν και σε άλλους τη συγκεκριμένη επιχείρηση. Από την άλλη, οι επικριτές είναι κατά κανόνα δυσαρεστημένοι πελάτες, που εκπέμπουν αρνητικά μηνύματα για τα προϊόντα ή τις υπηρεσίες που λαμβάνουν και συμβάλλουν στη διαμόρφωση αρνητικής εικόνας για την επιχείρηση (Gamon et al., 2008; Xu et al., 2011).

Πολλές ωστόσο φορές οι πληροφορίες που λαμβάνονται από τους πελάτες ή το κοινό έχουν τη μορφή μη δομημένων δεδομένων (unstructured data), γεγονός που καθιστά τη μαζική ανάλυσή τους μη τετριμμένη και απαιτητική. Στην περίπτωση αυτή, όπως έχει ήδη αναφερθεί, η ανάλυση συναισθημάτων πραγματοποιείται με τη χρήση ειδικών τεχνικών, όπως οι τεχνικές μηχανικής μάθησης (machine learning), με τη βοήθεια των οποίων ακόμη και τα μη δομημένα δεδομένα μπορούν να ερμηνευτούν και να αξιολογηθούν ως προς τα υποκείμενα συστήματα, αυτοματοποιημένα, με συνολικό και συστηματικό τρόπο. Με τη χρήση τεχνικών ανάλυσης συναισθημάτων μπορεί να πραγματοποιηθεί συγκέντρωση, επεξεργασία, ομαδοποίηση και ανάλυση όλων των απαντήσεων των πελατών, ώστε να προκύψει μια συνολική εικόνα που απεικονίζει την πολικότητα των συναισθημάτων και το βαθμό αφοσίωσης των πελατών. Τα όποια ευρήματα προκύψουν μπορούν να χρησιμοποιηθούν για αυτό που συχνά αποκαλείται βελτίωση της εμπειρίας του πελάτη (customer experience) [Kucher et al., 2018; Xu et al., 2011].

2.4.3 Παρακολούθηση συναισθημάτων κοινού σε βάθος χρόνου

Η διαχρονική εξέταση της εξέλιξης των συναισθημάτων του κοινού είναι κάτι που πολλές επιχειρήσεις εφαρμόζουν για να βλέπουν πως εξελίσσεται στο χρόνο ο τρόπος με τον οποίο αισθάνονται οι πελάτες τους. Ειδικότερα δε, αυτό που έχει ιδιαίτερη αξία και επιχειρείται πλέον σε πολλές περιπτώσεις είναι η όποια ανάλυση να πραγματοποιείται σε χρονικά σημεία που συντρέχουν ιδιαίτεροι λόγοι, όταν βρίσκεται για παράδειγμα σε εξέλιξη κάποια διαφημιστική καμπάνια ή όταν υπάρχουν σημαντικές ανακοινώσεις ή γεγονότα γύρω από το όνομα της επιχείρησης.

Θα μπορούσε εδώ να αναφερθεί το χαρακτηριστικό παράδειγμα της Nike, η οποία το Σεπτέμβριο του 2019 ανακοίνωσε ότι θα περιλάμβανε τον αμφιλεγόμενο ποδοσφαιριστή του Εθνικού Πρωταθλήματος Ποδοσφαίρου Colin Kaepernick, ο οποίος γονάτισε κατά την ανάκρουση του Εθνικού Ύμνου ως ένδειξη διαμαρτυρίας κατά της αστυνομικής βίας, στην καμπάνια εορτασμού συμπλήρωσης 30 ετών του εμβληματικού σήματος της Nike. Μεγάλος αριθμός εκνευρισμένων πελατών της εταιρίας κατέγραψαν τη δυσφορία τους στο Twitter, γεγονός που καταγράφηκε και από τη σχετική ανάλυση συναισθημάτων. Το ιδιαίτερα ενδιαφέρον εύρημα που ακολούθησε ήταν ότι, παρά την έντονα αρνητική δημοσιότητα, οι πωλήσεις της εταιρίας εκτοξεύτηκαν κατά 31% κατά την ίδια χρονική περίοδο, ενώ οι αναφορές του brand name της Nike στα μέσα κοινωνικής δικτύωσης σημείωσαν τρομακτική άνοδο κατά 1678% και αυτές του Kaepernick αυξήθηκαν κατά το εξωπραγματικό ποσοστό του 362.280%. Τα συγκεκριμένα στοιχεία εντοπίστηκαν μέσω της σχετικής ανάλυσης συναισθημάτων που πραγματοποιήθηκε. Γίνεται λοιπόν εύκολα αντιληπτό πως χωρίς την εν λόγω ανάλυση τα συμπεράσματα για τον αντίκτυπο της καμπάνιας πιθανότατα θα ήταν ελλιπή και παραπλανητικά (Wonderflow, 2018).

2.4.4 Ανάλυση αναφορών επωνυμίας επιχειρήσεων

Η παρακολούθηση των μέσων κοινωνικής δικτύωσης περιορίζεται στην εμβέλεια των ιστότοπων και εφαρμογών κοινωνικής δικτύωσης, αν και οι αναφορές της επωνυμίας μιας επιχείρησης (brand name) ή ενός οργανισμού εμφανίζονται και σε άλλα σημεία του διαδικτύου, όπως sites ειδήσεων, ιστολόγια, δελτία τύπου, κ.λπ. Στη συγκεκριμένη εφαρμογή ανάλυσης συναισθημάτων το πρώτο βήμα είναι η αναζήτηση και συλλογή αναφορών επωνυμίας επιχειρήσεων από διαδικτυακούς ιστότοπους και εφαρμογές κοινωνικής δικτύωσης και εν συνεχεία ακολουθεί η επεξεργασία και ανάλυση των δεδομένων που συλλέχθηκαν. Μία από τις

τεχνικές που χρησιμοποιείται αποτελεσματικά στη συγκεκριμένη περίπτωση είναι η τεχνική ανάλυσης κειμένων (text analytics), ώστε να ληφθούν υπόψη εκτός από τη συχνότητα εμφάνισης του brand name και τα συμφραζόμενα των εν λόγω εμφανίσεων (Liu et al., 2016). Το σχήμα που ακολουθεί είναι διευκρινιστικό της ανάλυσης αναφορών επωνυμίας.

Η ανάλυση συναισθημάτων, ώστε ένας οργανισμός ή μια επιχείρηση να κατανοήσει τα συναισθήματα που εκφράζονται σχετικά με την επωνυμία, επιτρέπει την κατανόηση και τον προσδιορισμό του πως αισθάνεται το κοινό για το συγκεκριμένο brand name. Τα οφέλη μιας τέτοιας ανάλυσης συνοψίζονται στα ακόλουθα (Liu et al., 2016):

- Μπορούν να ενεργοποιηθούν μηχανισμοί προειδοποίησης ανάλογα με το εύρος και το λόγο για τον οποίο εμφανίζεται ένα brand name στο διαδίκτυο.
- Μπορεί να προσδιοριστεί η πολικότητα του συνόλου του κοινού ή υποτμημάτων αυτού σχετικά με το brand name.
- Παρέχεται η δυνατότητα συνολικής άποψης για τη δημόσια αντίληψη που διαμορφώνεται για ένα brand name.

2.4.5 Διεξαγωγή ερευνών αγοράς

Εν γένει, κάθε διαδικτυακή αναζήτηση, που ο καθένας από εμάς πραγματοποιεί, θα μπορούσε να θεωρηθεί ως ένας απλός και γρήγορος τρόπος για την πραγματοποίηση έρευνας αγοράς. Το πρόβλημα αυτού του τύπου έρευνας είναι ότι σε ένα μεγάλο βαθμό τα αποτελέσματα αυτών των αναζητήσεων επηρεάζονται από τους αλγόριθμους αναζήτησης και τις εκάστοτε προτιμήσεις των μηχανών αναζήτησης. Εάν θέλει κανείς να πραγματοποιήσει πληρέστερη, αντικειμενικότερη και καλύτερα τεκμηριωμένη έρευνα αγοράς, τότε μπορεί να χρησιμοποιήσει αλγόριθμους και τεχνικές ανάλυσης συναισθημάτων. Τα δεδομένα πάνω στα οποία θα πραγματοποιηθεί η ανάλυση μπορούν να αντληθούν με την ευρεία σάρωση δεδομένων του διαδικτύου που σχετίζονται με το θέμα του ενδιαφέροντος, χρησιμοποιώντας τις κατάλληλες τεχνικές (Pang & Lee, 2008).

2.5 Τεχνικές οπτικοποίησης ανάλυσης συναισθημάτων

2.5.1 Εισαγωγή στις τεχνικές οπτικοποίησης συναισθημάτων

Ο όρος «οπτικοποίηση συναισθημάτων» (Sentiment Visualization) αναφέρεται στους τρόπους και τις τεχνικές, μέσα από τις οποίες μπορούν να απεικονιστούν και να αναλυθούν οι πληροφορίες ανάλυσης συναισθημάτων, που προέκυψαν από τη συλλογή και επεξεργασία δεδομένων κειμένου. Η οπτικοποίηση συναισθημάτων αποτελεί υποσύνολο της ευρύτερης ερευνητικής περιοχής που είναι γνωστή ως οπτικοποίηση κειμένου (text visualization) [Kucher & Kerren, 2015].

Η οπτικοποίηση των αναλύσεων συναισθημάτων και των στοιχείων που προκύπτουν από την επεξεργασία των σχετικών κειμένων αποτελεί ένα σύγχρονο πεδίο ερευνητικής δραστηριότητας εδώ και περισσότερο από μία δεκαετία. Τα αποτελέσματα αναλύσεων συναισθημάτων οπτικοποιούνται πλέον με πολλούς και διαφορετικούς τρόπους, από παραδοσιακά γραφήματα πίτας ή ραβδογράμματα μέχρι εξειδικευμένα γραφήματα οπτικοποίησης, που περιλαμβάνουν νέους τρόπους αναπαράστασης κατάλληλους για την ανάλυση πολύπλοκων και πολυδιάστατων συνόλων δεδομένων, που συνδυάζουν μεταξύ άλλων τις διαστάσεις χρόνου και χώρου και μπορούν την ίδια στιγμή να περιγράψουν αλληλοσυσχετίσεις (Kucher et al., 2018).

Η οπτικοποίηση συναισθημάτων ως περιοχή εμπεριέχει διαφορετικά θεματικά πεδία, όπως είναι η παρακολούθηση της κοινής γνώμης στα μέσα κοινωνικής δικτύωσης, η ψηφιακή ανάλυση βιβλιογραφίας ανθρωπιστικών επιστημών, η έρευνα ανάλυσης συναισθημάτων, η γλωσσολογία και η επεξεργασία φυσικής γλώσσας (Natural Language Processing). Κατά τις πρώτες φάσεις εξέλιξης του συγκεκριμένου ερευνητικού πεδίου η οπτικοποίηση συναισθημάτων συσχετίστηκε με την εξόρυξη δεδομένων (data mining) και την επεξεργασία φυσικής γλώσσας, ενώ οι τεχνικές οπτικοποίησης ήταν κατά κανόνα απλές. Πλέον, οι τεχνικές που σήμερα χρησιμοποιούνται αντανακλούν την πρόοδο που έχει σημειωθεί τα τελευταία χρόνια στο ευρύτερο πεδίο της οπτικοποίησης, είναι πολύ πιο σύνθετες και διαχειρίζονται ευρεία γκάμα ετερογενών δεδομένων. Οι γενικές ιδιότητες και προκλήσεις της οπτικοποίησης συναισθημάτων μπορούν να συνοψιστούν στα κάτωθι (Kucher et al., 2018):

- οι τεχνικές οπτικοποίησης συναισθημάτων χρησιμοποιούν δεδομένα, που είναι «δεμένα» με το εκάστοτε χρησιμοποιούμενο μοντέλο ανάλυσης συναισθημάτων και πεδίο εφαρμογής.

- η οπτικοποίηση συναισθημάτων καλύπτει μία ευρεία γκάμα θεμάτων, που κινούνται από την ανίχνευση υποκειμενικότητας έως την ανάλυση συναισθήματος και στάσης.
- Η οπτικοποίηση συναισθημάτων αντανακλά το μεγάλο εύρος τύπων δεδομένων και αναγκών των χρηστών που υπάρχουν στο εν λόγω ερευνητικό πεδίο και τις εφαρμογές ανάλυσης συναισθημάτων (π.χ. γλωσσολογία και επεξεργασία φυσικής γλώσσας, εφαρμογές κοινωνικής δικτύωσης, κ.λπ.), κάνοντας χρήση πολλαπλών οπτικών καναλιών και αναπαραστάσεων.

Η οπτικοποίηση συναισθημάτων δεν έχει προσελκύσει μέχρι σήμερα αντίστοιχο ερευνητικό ενδιαφέρον με άλλους τομείς οπτικοποίησης, που σχετίζονται επίσης με δεδομένα που εξάγονται από κείμενο, όπως είναι για παράδειγμα η οπτικοποίηση των θεματικών μοντέλων (topic models) ή γεγονότων (events). Στις λίγες μελέτες που έχουν διενεργηθεί επί του συγκεκριμένου αντικειμένου και συγκεντρώνουν την ερευνητική δραστηριότητα στο εν λόγω θεματικό πεδίο περιλαμβάνονται αυτές των Kucher και συν. (2018), Wanner και συν. (2014), Boumaïza (2015) και Shamim και συν. (2015). Η εξαγωγή και απεικόνιση της πολικότητας συναισθημάτων αποτελεί μία από τις πλέον συνήθεις μεθόδους οπτικοποίησης που επιλέγονται (Kucher et al., 2018). Στο έργο του Boumaïza (2015) αναφέρονται 35 τεχνικές οπτικοποίησης συναισθηματικών κριτικών, με έμφαση στα κείμενα των μέσων κοινωνικής δικτύωσης, ενώ στη μελέτη των Shamim και συν. (2015) πραγματοποιείται μια ανασκόπηση 11 τεχνικών, για τις οποίες διεξάγεται μια συγκριτική ανάλυση ως προς διάφορα χαρακτηριστικά τους, όπως είναι η φιλικότητα προς το χρήστη και ο βαθμός χρησιμότητας. Τέλος, στην έρευνα των Kucher και συν. (2018) εξετάζονται 132 τεχνικές οπτικοποίησης και κατηγοριοποιούνται βάσει κριτηρίων, που σχετίζονται με το υπολογιστικό μοντέλο, τα δεδομένα, τις διαδικασίες των χρηστών και την οπτική αναπαράσταση.

Στην ενότητα που ακολουθεί παρουσιάζεται μια λεπτομερής ανασκόπηση και κατηγοριοποίηση των τεχνικών οπτικοποίησης συναισθημάτων, ομαδοποιημένων σε κατηγορίες ανάλογα με τα χαρακτηριστικά των χρησιμοποιούμενων δεδομένων, τις πηγές προέλευσης και ιδιότητες των δεδομένων, τις εργασίες που εκτελούνται και τις παραμέτρους και μεταβλητές οπτικοποίησης.

2.5.2 Τεχνικές με βάση τα χαρακτηριστικά των δεδομένων

Η πρώτη από τις κατηγοριοποιήσεις, που παρατίθενται στην ενότητα αυτή, προκύπτει με βάση τα επιμέρους χαρακτηριστικά των χρησιμοποιούμενων δεδομένων, περιλαμβανομένου του θεματικού πεδίου ανάλυσης και της αναπαράστασης της πηγής δεδομένων. Οι περισσότερες τεχνικές οπτικοποίησης συναισθημάτων έχουν δημιουργηθεί πάνω σε συγκεκριμένες περιοχές δεδομένων (data domains), κάποιες από τις οποίες έχουν διαχρονικά συσχετιστεί με τις διαδικασίες ανάλυσης συναισθημάτων και επεξεργασίας φυσικής γλώσσας. Χαρακτηριστικό παράδειγμα ενός τέτοιου τομέα, που παρέχει πολλά δεδομένα κειμένου κατάλληλα για ανάλυση συναισθημάτων, είναι τα διαδικτυακά μέσα κοινωνικής δικτύωσης, συμπεριλαμβανομένων forums, ιστολογίων και μέσων κοινωνικής δικτύωσης (Kucher et al., 2018).

Από τις πρώτες εργασίες που δημοσιεύτηκαν στο συγκεκριμένο τομέα ήταν των Mishne και De Rijke (2006), η οποία περιέγραφε ένα σύστημα επεξεργασίας φυσικής γλώσσας (MoodViews) για την ανάλυση της διάθεσης των χρηστών ιστολογίων. Το σύστημα χρησιμοποιούσε ετικέτες (tags) διάθεσης, που εισάγονταν από τους χρήστες, πραγματοποιούσε όμως και προβλέψεις για πολλούς τύπους διάθεσης που προτεινονταν από το ιστολόγιο. Ως προς τη γραφική αναπαράσταση, εφαρμόζονταν βασικές τεχνικές απεικόνισης για χρονική οπτικοποίηση, ενώ έδινε τη δυνατότητα στους χρήστες να διερευνούν σε λεπτομέρεια διαδραστικά τους σημαντικούς όρους και φράσεις για επιλεγμένα χρονικά διαστήματα. Στην έρευνα των Abbasi και Chen (2007) παρουσιάστηκε μια τεχνική εξερεύνησης και ανάλυσης εγγράφων, που χρησιμοποιούσε μεταφερόμενες φυσαλίδες (bubbles) για την επισήμανση περιοχών ενδιαφέροντος σε κείμενα επικοινωνίας, ιστολόγια και μέσα κοινωνικής δικτύωσης.

Με την ραγδαία ανάπτυξη των μικροιστολογίων (microblogs) που ακολούθησε, όλο και περισσότερες τεχνικές άρχισαν να εστιάζουν σε αυτού του τύπου τα δεδομένα. Στη μελέτη των Diakopoulos και συν. (2010) περιγράφεται το VoxCivitas, ένα σύστημα οπτικής ανάλυσης για τη διερεύνηση συναισθημάτων, συνάφειας και λέξεων-κλειδιών που σχετίζονται με δημόσιες εκδηλώσεις σε αναρτήσεις μέσω κοινωνικής δικτύωσης. Το συγκεκριμένο σύστημα χρησιμοποιεί ως γραφική αναπαράσταση ένα ραβδόγραμμα στοίβας (stacked bar) για την αναπαράσταση των τιμών. Στην έρευνα των Marcus και συν. (2011) παρουσιάστηκε ένα σύστημα οπτικής ανάλυσης δημοσιεύσεων στο Twitter, που υποστήριζε την ανίχνευση πολικότητας συναισθημάτων. Η αθροιστική τιμή πολικότητας απεικονίζονταν μέσω ενός γραφήματος πίτας

(pie chart), ενώ η αναπαράσταση πολικότητας για μεμονωμένα έγγραφα γίνονταν με την χρωματική επισήμανση γεωγραφικού χάρτη. Παρόμοιες δουλειές που εντάσσονται στη συγκεκριμένη κατηγορία είναι τα συστήματα BLEWS (Gamon et al., 2008), ForAVis (Wanner et al., 2011), webLyzard (Scharl et al., 2012), MoodLens (Zhao et al., 2012), Whisper (Cao et al., 2012), Opinion-Flow (Wu et al., 2014), PEARL (Zhao et al., 2014b), SentiCompass (Wang et al., 2015), uVSAT (Kucher et al., 2016) και πολλά ακόμη.

Σημαντικός αριθμός ερευνητικών εργασιών έχουν ασχοληθεί με την ανάλυση και οπτικοποίηση συναισθημάτων σε κείμενα επικοινωνίας όπως τα μηνύματα ηλεκτρονικού ταχυδρομείου (emails) και οι διαδικτυακές συζητήσεις (chats). Ένα από τα πρώτα παραδείγματα δουλειάς στην περιοχή αυτή είναι το CrystalChat (Tat & Carpendale, 2006), μια τεχνική οπτικοποίησης για το ιστορικό προσωπικών συνομιλιών, που αναπαριστά τα μηνύματα ως κύκλους και τα ομαδοποιεί σε μια τρισδιάστατη απεικόνιση, ώστε να απεικονίσει ταυτόχρονα και πληροφορίες χρονικής ακολουθίας και των ατόμων της συνομιλίας. Η συγκεκριμένη τεχνική κωδικοποιεί και απεικονίζει το συναισθηματικό περιεχόμενο των συνομιλιών, λαμβάνοντας υπόψη τα emoticons που χρησιμοποιούνται εντός των διαλόγων. Στην έρευνα των Gobron και συν. (2010) παρουσιάζεται μια ιδιαίτερη τεχνική οπτικοποίησης συναισθημάτων τέτοιου τύπου μηνυμάτων, όπου χρησιμοποιείται το Facial Action Coding System (FACS) για τη δημιουργία τρισδιάστατων κινούμενων ειδώλων, των οποίων τα πρόσωπα αντανakλούν τα συναισθήματα που σχετίζονται με τα αντίστοιχα δεδομένα κειμένου. Στο έργο των Chen και συν. (2014) διεξάγεται επίσης μια οπτική ανάλυση και απεικόνιση των συνομιλιών, χρησιμοποιώντας emoticons που διαβαθμίζονται ανάλογα με την πολικότητα. Άλλες τεχνικές οπτικοποίησης συναισθημάτων των μηνυμάτων ηλεκτρονικού ταχυδρομείου παρουσιάζονται στις μελέτες των Abbasi και Chen (2007), Mohammad (2012), Das και συν. (2012), Guzman (2013) και Cao και συν. (2016).

Πριν από την περίοδο που καθιερώθηκε η ευρεία χρήση των μέσων κοινωνικής δικτύωσης, η ανάλυση συναισθημάτων χρησιμοποιήθηκε σχεδόν αποκλειστικά για την αξιολόγηση προϊόντων και σχολίων των πελατών επιχειρήσεων. Επομένως, υφίσταται μια διόλου αμελητέα κατηγορία ανάλυσης και οπτικοποίησης συναισθημάτων επί κριτικών/ αναφορών, η οποία περιλαμβάνει δουλειές όπως το «*Affect Inspector*» (Subasic & Huettner, 2001), που χρησιμοποιεί αστεράκια για την οπτικοποίηση των κριτικών κειμένου και ταινιών, το Pulse (Gamon et al., 2005), που χρησιμοποιεί έναν χάρτη δέντρων για να απεικονίσει ομάδες συναισθημάτων για κριτικές αυτοκινήτων και το Opinion Observer (Liu et al., 2005), όπου η οπτικοποίηση των

απόψεων των πελατών πραγματοποιείται με τη χρήση ραβδογραφημάτων. Το AMAZING (Miao et al., 2009) οπτικοποιεί κριτικές προϊόντων, που απεικονίζει τα αποτελέσματα της ανάλυσης συναισθημάτων μέσω γραφημάτων γραμμής (line chart) και πίτας (pie chart), ενώ στο έργο των Chen και συν. (2006) γίνεται χρήση πολλαπλών τεχνικών ανάλυσης και οπτικοποίησης για τη διερεύνηση και απεικόνιση αντιφατικών απόψεων κριτικών πελατών. Στη μελέτη των Hao και συν. (2013) παρουσιάζονται τεχνικές οπτικοποίησης για την ανάλυση σχολίων πελατών, όπως είναι ο γεωγραφικός χάρτης απεικόνισης σημείων (pixel sentiment geo map), γεωγραφικός χάρτης απεικόνισης βασικών όρων (key term geo map), το ημερολόγιο συναισθημάτων επί σημείων (pixel based sentiment calendar) και χάρτης αυτοσυσχετιζόμενων όρων (self-organizing term association map). Τέλος, στην έρευνα των Oelke και συν. (2009) παρουσιάζονται τεχνικές οπτικοποίησης όπως ο χάρτης συστάδων απόψεων (map of opinion clusters), ενώ στη μελέτη των Xu και συν. (2011) τα αποτελέσματα της ανάλυσης απεικονίζονται σε συγκριτικούς χάρτες σχέσεων (comparative relation maps). Ως γενικό σχόλιο, πολλές από τις τεχνικές που χρησιμοποιούνται στη συγκεκριμένη κατηγορία, προέρχονται από την επιστημονική περιοχή της επεξεργασίας φυσικής γλώσσας και εστιάζουν κυρίως στο αναλυτικό μέρος, χρησιμοποιώντας για την οπτικοποίηση απλές κατά κανόνα γραφικές αναπαραστάσεις.

Εφαρμογές ανάλυσης και οπτικοποίησης συναισθημάτων υπάρχουν και στη λογοτεχνία, αν και σε περιορισμένη κλίμακα. Στην έρευνα των Subasic και Huettner (2001) προτάθηκε ο Επιθεωρητής Επιρροής (Affect Inspector) για την ανάλυση συναισθημάτων σε ποιήματα. Οι Liu και συν. (2003) χρησιμοποιούν μία χρωματισμένη ράβδο για να οπτικοποιήσουν τα συναισθήματα που σχετίζονται με τμήματα λογοτεχνικών κειμένων, ενώ στην έρευνα του Mohammad (2012) παρουσιάζεται η ανάλυση συναισθημάτων με τεχνικές λεξικής αντιστοίχισης (lexical matching) όρων, που σχετίζονται με οκτώ τύπους συναισθημάτων και θετική ή αρνητική κατηγοριοποίηση σε κείμενα αλληλογραφίας και βιβλία καθώς επίσης και τεχνικές οπτικοποίησης όπως γραφήματα γραμμών, ράβδων και σύννεφα λέξεων (word clouds). Ακόμη, στην έρευνα των Weiler και συν. (2015) παρουσιάζεται το «*Stor-e-Motion*», ένα σύστημα ανάλυσης και οπτικοποίησης ροών κειμένου με γραφική αναπαράσταση των συναισθημάτων ως γράφημα ροής (stream graph).

Στην περιοχή της οπτικοποίησης συναισθημάτων επιστημονικών άρθρων υπάρχει μικρής έκτασης ερευνητική δραστηριότητα, η οποία κυρίως σχετίζεται με την ανάλυση και οπτικοποίηση των παραπομπών (citations). Στην έρευνα των Schäfer και Spurk (2010) εξετάζονται και κατηγοριοποιούνται βάσει πολικότητας οι παραπομπές σε επιστημονικά άρθρα, ενώ τα

αποτελέσματα της ανάλυσης οπτικοποιούνται μέσω χρωματικής απεικόνισης του γράφου παραπομπών. Στο Small (2011) παρουσιάζεται η ανάλυση συναισθημάτων και οπτικοποίηση των βιβλιογραφικών παραπομπών με Χάρτες της Επιστήμης (Maps of Science). Στη συγκεκριμένη μελέτη, πέραν της συνήθους κατηγοριοποίησης συναισθημάτων βάσει πολικότητας, παρουσιάζεται και ανάλυση επί της αβεβαιότητας και διαφοροποίησης/ αντίθεσης, ενώ η οπτικοποίηση γίνεται μέσω γραφήματος κόμβων και διασυνδέσεων. Τέλος, στην έρευνα των Wang και συν. (2016) παρουσιάζεται η ανάλυση και κατηγοριοποίηση συναισθημάτων βάσει πολικότητας και η οπτικοποίηση των αποτελεσμάτων μέσω ενός γράφου βιβλιογραφικών παραπομπών.

Μία ακόμη περιοχή εφαρμογής της ανάλυσης και οπτικοποίησης συναισθημάτων είναι αυτή των μέσων επικοινωνίας, όπως είναι οι ιστότοποι ειδήσεων, τα sites γνώσεων (π.χ. Wikipedia) και άλλα αντίστοιχα. Παράδειγμα ερευνητικής δραστηριότητας στην περιοχή αυτή είναι το SATISFI (Sentiment and Time Series: Financial Analysis System), που παρουσιάστηκε στην έρευνα των Taskaya και Ahmad (2003), και στο οποίο χρησιμοποιείται λεξική αντιστοίχιση για την ανάλυση πολικότητας των οικονομικών ειδήσεων. Οι τιμές πολικότητας ομαδοποιούνται και οπτικοποιούνται με απλά γραφήματα. Σε άλλη περίπτωση, οι Fukuhara και συν. (2007) χρησιμοποίησαν απλά γραφήματα γραμμών για την οπτικοποίηση συναισθημάτων που σχετίζονται με οκτώ επιμέρους κατηγορίες σε κείμενα ειδήσεων και ιστολογίων. Ένα ακόμη παράδειγμα είναι το σύστημα BLEWS (Gamon et al., 2008), το οποίο εστιάζει στην ανάλυση των σχέσεων μεταξύ άρθρων ειδήσεων και πολιτικών δημοσιεύσεων σε ιστολόγια. Η οπτικοποίηση της ανάλυσης πραγματοποιείται στην περίπτωση αυτή με οπτική επισήμανση λάμψης γύρω από ραβδογράμματα που αναπαριστούν τις επιμέρους κατηγορίες της ανάλυσης. Στην έρευνα των Zhang και συν. (2009) παρουσιάστηκε ως μέσο απεικόνισης ο χάρτης συναισθημάτων (Sentiment Map), στον οποίο απεικονίζονται οι κατηγορίες συναισθημάτων που έχουν ανιχνευθεί σε άρθρα ειδήσεων, λαμβάνοντας υπόψη και τη χρονική διάσταση. Στη μελέτη των Cui και συν. (2012) παρουσιάζεται το TextWheel, το οποίο εισάγει περισσότερο σύνθετες τεχνικές οπτικοποίησης, συνδυάζοντας πολλαπλές γραφικές απεικονίσεις, όπως το γράφημα ακτίνας κόμβου και σχέσεων (radial node link), μια ζώνη μεταφοράς σχήματος «U» και ένα γράφημα τάσεων σημαντικότητας.

Εκτός όλων των παραπάνω, υπάρχει ένας αριθμός τεχνικών οπτικοποίησης συναισθημάτων, οι οποίες δε μπορούν να καταταγούν σε κάποια από τις προαναφερθείσες κατηγορίες, αφού δεν αναφέρονται σε κάποιο συγκεκριμένο τομέα δεδομένων. Μία από αυτές τις

τεχνικές παρουσιάζεται από τους Duan και συν. (2012) και είναι το Σύστημα VISA, το οποίο χρησιμοποιεί ένα γράφημα ροής (stream graph) με ενσωματωμένα σύννεφα ετικετών (tag clouds), και πρόσθετες συμπληρωματικές απεικονίσεις όπως γραφήματα πίτας, ραβδογραφήματα, κ.λπ. Μία ακόμη τεχνική παρουσιάζεται στη μελέτη των Wecker και συν. (2014) και χρησιμοποιεί για οπτικοποίηση συγκεκριμένες γραμματοσειρές και χρώμα φόντου. Οι Neviarouskaya και συν. (2014) οπτικοποιούν τα αποτελέσματα της ανάλυσης εγγράφων με το μοντέλο «@AM», το οποίο απεικονίζει και στοιχεία ανάλυσης πέραν της πολικότητας, όπως είναι η κρίση και η εκτίμηση. Οι Gold και συν. (2015) οπτικοποιούν τα συναισθήματα, κάνοντας χρήση ειδικής σήμανσης συγκεκριμένων λέξεων και τμημάτων ράβδων.

2.5.3 Τεχνικές με βάση την πηγή προέλευσης και ιδιότητες των δεδομένων

Όπως εύκολα μπορεί να γίνει κατανοητό, ο τύπος του αρχείου προέλευσης δεδομένων κειμένου έχει επίπτωση στη σχεδίαση τεχνικών απεικόνισης συναισθημάτων. Βέβαια, σε περιπτώσεις μεγάλων εγγράφων, μπορεί να μην είναι ξεκάθαρη η εφαρμογή συγκεκριμένων τεχνικών λόγω της ενδεχόμενης ποικιλομορφίας του περιεχομένου. Αντίστροφα, σε άλλες περιπτώσεις μπορεί να απαιτείται η ενιαία ανάλυση και οπτικοποίηση συναισθημάτων πολλαπλών κειμένων, οπότε και πάλι μπορεί να μην είναι σαφές ποιες τεχνικές είναι εφαρμόσιμες. Τρεις είναι οι βασικές πηγές κειμένων, στις οποίες πραγματοποιείται ανάλυση και οπτικοποίηση συναισθημάτων: τα αρχεία κειμένου (documents), οι συλλογές εργασιών (corpora) και οι ροές κειμένου (streams) [Kucher et al., 2018]. Ακολουθώς αναλύονται οι τρεις αυτές κατηγορίες.

Καταρχάς, στις περιπτώσεις μεμονωμένων αρχείων κειμένου, οι τεχνικές οπτικοποίησης της ανάλυσης συναισθημάτων έχουν μελετηθεί από αρκετούς ερευνητές. Στη μελέτη των Subasic και Huettner (2001) παρουσιάζονται τεχνικές απλής ή πολλαπλής ταυτόχρονα οπτικοποίησης εγγράφων, ενώ το προταθέν Σύστημα «uVSAT» (Kucher et al., 2016) παραθέτει τεχνικές οπτικοποίησης μεμονωμένων εγγράφων με βάση επιλεγμένα υποσύνολα συγκεντρωτικών σειρών τιμών συναισθημάτων. Στην έρευνα των Ruppert και συν. (2016) παρουσιάζονται οπτικοποιημένες περιλήψεις αναλύσεων υποκειμενικότητας και πολικότητας για μεμονωμένα έγγραφα, ενώ οι Gobron και συν. (2010) και Krcadinac και συν. (2016) οπτικοποιούν πραγματικού χρόνου ροές κειμένου, χρησιμοποιώντας τεχνικές αρχείων κειμένου. Άλλες ερευνητικές εργασίες που παρουσιάζουν τεχνικές οπτικοποίησης συναισθημάτων σε μεμονωμένα έγγραφα είναι αυτές

των Liu και συν. (2003), Wecker και συν. (2014), Gold και συν. (2015), Li και Ren (2009), Almutairi (2013) και Bembenik και Andruszkiewicz (2016).

Ευρεία γκάμα τεχνικών ανάλυσης συναισθημάτων υποστηρίζουν την οπτικοποίηση δεδομένων κειμένου, που ανήκουν σε συλλογές εγγράφων, γνωστές και ως corpus ή corpora. Σε αντίθεση με την απεικόνιση μεμονωμένων εγγράφων, οι τεχνικές οπτικοποίησης που υποστηρίζουν συλλογές εγγράφων, έχουν να αντιμετωπίσουν προκλήσεις όπως μεγάλους όγκους δεδομένων, ετερογένεια κειμένου, συσχετίσεις μεταξύ εγγράφων, κ.λπ. Χαρακτηριστικό παράδειγμα αυτής της κατηγορίας αποτελούν οι τεχνικές που παρουσιάζονται στις μελέτες των Xu και συν. (2011) και Wu και συν. (2014) και οι οποίες οπτικοποιούν συναισθήματα, που προκύπτουν από την ανάλυση συλλογής σχολίων πελατών ή ροών στο twitter. Άλλες τεχνικές αυτής της κατηγορίας παρουσιάζονται στις μελέτες των Subasic και Huettner (2001), Liu και συν. (2005), Gamon και συν. (2005), Wu και συν. (2010), Diakopoulos και συν. (2010), Hoque και Carenini (2014), Cao και συν. (2015) και Mohammad και συν. (2016).

Τα τελευταία χρόνια έχουν πλέον κάνει έντονα αισθητή την παρουσία τους στη βιβλιογραφία τεχνικές οπτικοποίησης ανάλυσης συναισθημάτων πάνω σε ροές δεδομένων (streams). Μία από τις συνηθέστερες πηγές άντλησης δεδομένων των τεχνικών αυτών είναι τα microblogs (π.χ. Twitter, Weibo) [Marcus et al., 2011; Zhao et al., 2012]. Από τις δημοφιλέστερες μελέτες αυτής της κατηγορίας είναι αυτές των Liu και συν. (2013), Kranjc και συν. (2015), Steed και συν. (2015) και Tsirakis και συν. (2016).

Πέραν της πηγής προέλευσης των δεδομένων, η έρευνα στην περιοχή της οπτικοποίησης συναισθημάτων διαχωρίζει πολλές φορές τις τεχνικές και με βάση τις ιδιότητες των δεδομένων που χρησιμοποιούνται. Υπό αυτή την οπτική, μία από τις πολύ ενδιαφέρουσες κατηγορίες είναι αυτή που περιέχει τεχνικές που κάνουν χρήση γεωχωρικών πληροφοριών (geospatial information). Στην έρευνα των Zhang και συν. (2009) η οπτικοποίηση των συναισθημάτων πραγματοποιείται μέσω γραφημάτων που συνδέουν τα αποτελέσματα της ανάλυσης με αντίστοιχες γεωγραφικές περιοχές. Σε αντίστοιχη λογική, στο (Scharl et al., 2012) παρουσιάζεται μια πλατφόρμα παρακολούθησης και οπτικής ανάλυσης των ειδήσεων που κοινοποιούνται σε μέσα κοινωνικής δικτύωσης, μαζικής ενημέρωσης, και άλλες πηγές του διαδικτύου. Τέλος στις μελέτες των Zhang και συν. (2013) και Caragea και συν. (2014) παρουσιάζονται τεχνικές αυτού του τύπου οι οποίες

συνδυάζουν τη χρήση ενός κοινού γεωγραφικού χάρτη με τις αναπαραστάσεις της ανάλυσης συναισθημάτων για συγκεκριμένες πόλεις ή περιοχές αντιπροσωπεύουν το δίκτυο συζητήσεων.

Μεγάλος αριθμός τεχνικών ανάλυσης συναισθημάτων στηρίζεται στην ανάλυση χρονικών ακολουθιών δεδομένων ή χρονοσειρών (time series). Οι περισσότερες από αυτές τις τεχνικές χρησιμοποιούν για την οπτικοποίηση γραφήματα απεικόνισης των συναισθημάτων στη διάσταση του χρόνου, από απλά διαγράμματα «γραμμής» (Mishne & De Rijke, 2006; Das et al., 2012; Hoerber et al., 2016), μέχρι γράφους ροών δεδομένων (Dual et al., 2012; Liu et al., 2013; Wu et al., 2014; Zhao et al., 2014b), γλύφους (glyph) [Wanner et al., 2009; Wanner et al., 2012; Lu et al., 2014] και σύνθετες κινούμενες αναπαραστάσεις (Gali et al., 2012; Zhang et al., 2013; El-Assady et al., 2016].

Μία ακόμη κατηγορία τεχνικών οπτικοποίησης συναισθημάτων είναι αυτές που χρησιμοποιούν δίκτυα (Networks) για την απεικόνιση εισερχόμενων ή υπό επεξεργασία δεδομένων. Χαρακτηριστικά παραδείγματα τέτοιων τεχνικών περιλαμβάνουν γραφήματα παραπομπών (citation graphs) (Schäfer & Spurk, 2010; Small, 2011; Wang, et al., 2016) και γραφήματα συσχετίσεων όρων, σχολίων, χρηστών, και κοινοτήτων (Bras-Oveanu et al., 2012; Chen et al., 2015; Wang et al., 2013; Cao et al., 2015). Άλλες ενδιαφέρουσες εργασίες που περιγράφουν τεχνικές που εμπίπτουν στη συγκεκριμένη κατηγορία, παρουσιάζονται στις έρευνες των Hoque και Carenini (2014), Hoque και Carenini (2016) και Prasojo και συν. (2015).

2.5.4 Τεχνικές με βάση τις εκτελούμενες εργασίες

Η παρούσα ενότητα εστιάζει στην οπτικοποίηση ανάλυσης συναισθημάτων με βάση τις εκτελούμενες εργασίες (tasks) στις χρησιμοποιούμενες τεχνικές. Η πρώτη κατηγορία τεχνικών υπό αυτή την οπτική είναι αυτές που χρησιμοποιούν εργασίες ανάλυσης και μπορούν να συνοψιστούν στις ακόλουθες (Kucher et al., 2018):

- 1) Η Ανάλυση Πολικότητας (Polarity Analysis) ή Ανίχνευση Υποκειμενικότητας (Subjectivity Detection) είναι η δημοφιλέστερη εργασία ανάλυσης που χρησιμοποιείται σε τεχνικές ανάλυσης συναισθημάτων. Οι τεχνικές που στηρίζονται σε τέτοιες εργασίες παρέχουν μια σύνοψη της πολικότητας κειμένων. Παραδείγματα τέτοιων τεχνικών είναι η SATISFI (Taskaya & Ahmad, 2003) και η Vox Civitas (Diakopoulos et al., 2010), ενώ οι

Annett και Kondrak (2008) παρουσίασαν το eNulog, ένα εργαλείο οπτικοποίησης συναισθημάτων σχολίων blogs για ταινίες, το οποίο διαχωρίζει τα σχόλια σε θετικά, ουδέτερα ή αρνητικά, χρωματίζοντας ανάλογα τις περιοχές των κειμένων. Στη μελέτη των Puri και συν. (2014) παρουσιάστηκε το σύστημα Ent-it-UP, το οποίο υπολόγιζε και απεικόνιζε την πολικότητα συναισθημάτων ατομικών όρων, ενώ στο έργο των Brooks και συν. (2014) παρουσιάστηκε το σύστημα Agave, που οπτικοποιεί με απλές γραφικές παραστάσεις την πολικότητα συναισθημάτων δεδομένων, που προέρχονται από μέσα κοινωνικής δικτύωσης. Τέλος, στο έργο των Gamon και συν. (2008) παρουσιάζεται το σύστημα BLEWS, το οποίο εστιάζει στην οπτικοποίηση των συσχετίσεων μεταξύ άρθρων ειδήσεων και δεδομένων πολιτικής από blogs, χρησιμοποιώντας ανάλυση υποκειμενικότητας και έχοντας ως βάση την οπτική αναπαράσταση σε ράβδους με πρόσθετα εφέ γύρω από αυτές ανάλογα με το βαθμό υποκειμενικότητας.

- 2) Η Μονοδιάστατη (Aspect based) Ανάλυση Συναισθήματος είναι επίσης δημοφιλής εργασία, που εμπεριέχεται σε τεχνικές ανάλυσης και οπτικοποίησης συναισθημάτων και η οποία, όπως έχει ήδη αναφερθεί, εστιάζει σε συγκεκριμένες πτυχές θεμάτων ή οντοτήτων. Η συγκεκριμένη κατηγορία ανάλυσης έχει συχνά χρησιμοποιηθεί πάνω σε δεδομένα κριτικών πελατών. Στο έργο των Yatani και συν. (2011) παρουσιάστηκε το Review Spotlight, ένα εργαλείο οπτικοποίησης των κριτικών πελατών μέσω νέφους ετικετών (tag cloud). Η πολικότητα των συναισθημάτων υπολογίζεται σε επίπεδο λέξης και απεικονίζεται με τη χρήση διαφορετικών χρωμάτων. Στο έργο των Alper και συν. (2011) παρουσιάζεται το OpinionBlocks, το οποίο χρησιμοποιεί για την οπτικοποίηση συναισθημάτων ανάλυσης σχολίων πελατών, συνδυασμό πολλαπλών ραβδογραμμμάτων και ετικετών κειμένου με δυνατότητα αλληλεπίδρασης με το χρήστη. Οι Di Caro και Grella (2013) παρουσίασαν το SentiVis, που οπτικοποιεί τα αποτελέσματα ανάλυσης συναισθημάτων βάσει συγκεκριμένων διαστάσεων κριτικής πελατών, χρησιμοποιώντας γραφήματα διασποράς (scatter) ή γραμμής (line). Στο έργο των Gorg και συν. (2013) παρουσιάζεται το σύστημα Jigsaw, το οποίο οπτικοποιεί την πολικότητα συναισθημάτων σε επίπεδο εγγράφου καθώς και σε επίπεδο επιμέρους διαστάσεων/ θεμάτων εγγράφων, χρησιμοποιώντας νέφη λέξεων (word clouds) και προβολές σε πλέγματα σημείων (pixel based). Οι Wensel και Sood (2008) παρουσιάζουν το σύστημά VIBES, το οποίο πραγματοποιεί ανάλυση συναισθημάτων συγκεκριμένων θεμάτων, που προέρχονται από

προσωπικές απόψεις που διατυπώνονται σε blogs και οπτικοποιεί την πολικότητα με χρήση γραφημάτων γραμμής (line plots), γλύφων (glyphs) και νεφών ετικετών (tag clouds). Τέλος, στις μελέτες των Cervantes και συν. (2015) και Mahmud και συν. (2016) παρουσιάζεται οπτικά η πολικότητα θεμάτων ενδιαφέροντος που συζητούνται στο twitter.

- 3) Ανάλυση Συναισθημάτων/ Επιπτώσεων (Emotion/ Affect Analysis) αποτελεί εργασία που ξεφεύγει από την ανάλυση πολικότητας και εστιάζει στη μοντελοποίηση και οπτικοποίηση των συναισθημάτων, κάνοντας χρήση κάποιου μοντέλου συναισθημάτων. Το σύστημα Affect Inspector (Subasic & Huettner, 2001) χρησιμοποιεί γραφήματα αστεριών (star plots) για την απεικόνιση συναισθημάτων εγγράφων κειμένου, ενώ ταυτόχρονα υποστηρίζει εξερεύνηση και επισήμανση των κειμένων. Στο έργο των Gregory και συν. (2006) παρουσιάζεται το σύστημα INSPIRE, που μεταξύ άλλων οπτικοποιεί τα συναισθήματα που προκύπτουν από την ανάλυση σχολίων πελατών, κάνοντας χρήση γραφήματος σε σχήμα τριαντάφυλλου (rose plot), στο οποίο απεικονίζονται στατιστικά για οκτώ κατηγορίες επιπτώσεων. Επίσης, σημαντική στη συγκεκριμένη κατηγορία είναι η δουλειά των Kang και Ren (2011), οι οποίοι οπτικοποιούν χρωματικά τα συναισθήματα που προκύπτουν από την ανάλυση θεμάτων σε δεδομένα blogs, κάνοντας χρήση γραφημάτων σύνδεσης κόμβων (node link diagrams). Στο έργο των Zhao και συν. (2014a) παρουσιάζεται το σύστημα PEARL, στο οποίο οπτικοποιούνται τα συναισθήματα που ανιχνεύονται σε tweets, λαμβάνοντας υπόψη τη χρονική διάσταση με την πάροδο του χρόνου. Το PEARL χρησιμοποιεί γραφήματα κινούμενων ροών, που μοιάζουν με ποταμούς (river-like streams) σε συνδυασμό με άλλες υποστηρικτικές αναπαραστάσεις όπως γραφήματα γραμμών (line graphs) και περιοχών (area charts), γλύφους (glyphs), γραφήματα διασποράς (scatter plots) και νέφη ετικετών (tag clouds). Τέλος, στο έργο των Wang και συν. (2015) παρουσιάζεται το σύστημα SentiCompass, το οποίο οπτικοποιεί συναισθήματα που ανιχνεύονται σε tweets, κάνοντας χρήση γραφικών, όπως γραφήματα αστεριών (star plots) που συνδυάζονται σε διάταξη σπείρας για να υποστηρίξουν την ταυτόχρονη οπτική ανάλυση πολλαπλών χρονικών διαστημάτων, και άλλες απλούστερες βοηθητικές αναπαραστάσεις.
- 4) Η Ανάλυση Στάσης (Stance Analysis) αποτελεί εργασία η οποία εστιάζει όχι στην κλασική ανάλυση συναισθήματος ή επιπτώσεων, αλλά και σε επιπλέον κατηγορίες αξιολόγησης που μπορούν να εκφραστούν σε ένα κείμενο, όπως είναι η βεβαιότητα

(certainty) και η κρίση (judgement). Παραδείγματα τέτοιων τεχνικών αποτελούν αυτά που έχουν παρουσιαστεί στις μελέτες των Small (2011), Almutairi (2013), Neviarouskaya και συν. (2014) και Bembenik και Andruszkiewicz (2016). Το σύστημα uVSAT (Kucher et al., 2016) αποτελεί επίσης παράδειγμα συστήματος, το οποίο οπτικοποιεί ανάλυση στάσης, η οποία ανιχνεύει την έκφραση των έξι συναισθημάτων του Ekman και επιπλέον την αβεβαιότητα και τη βεβαιότητα σε δεδομένα μέσω κοινωνικής δικτύωσης και τη χρονική τους εξέλιξη. Η γραφική απεικόνιση στηρίζεται σε αναπαραστάσεις γραφημάτων γραμμής (line plots), επισημασμένου κειμένου (highlighted text), διασποράς (scatter plots) και φυσαλίδων (bubble charts). Στο έργο των Mohammad και συν. (2016) εισάγεται η γραφική απεικόνιση δεδομένων στάσης μέσω dashboard, που εμπεριέχει γραφικές αναπαραστάσεις ραβδογραμμάτων και χάρτη-δέντρου (tree-map). Τέλος, οι El-Assady και συν. (2016) παρουσίασαν το σύστημα ConToVi, το οποίο επιτρέπει στους χρήστες να παρακολουθούν τη στάση των ομιλητών πάνω με συγκεκριμένα θέματα.

Η δεύτερη κατηγορία τεχνικών αυτού του τύπου είναι αυτές που στηρίζονται σε εργασίες αναπαράστασης και αλληλεπίδρασης, οι οποίες υποστηρίζονται απευθείας από τεχνικές απεικόνισης. Οι επιμέρους υποκατηγορίες της συγκεκριμένης κατηγορίας είναι οι ακόλουθες:

- 1) Η Αναπαράσταση Περιοχής Ενδιαφέροντος (Region of Interest) περιλαμβάνει τεχνικές που υποστηρίζουν την αυτόματη ανίχνευση και επισήμανση δεδομένων ενδιαφέροντος (Cui et al., 2012; Zhao et al., (2014a); Cao et al., 2016; Lu et al., 2016), περιοχών ενδιαφέροντος (Kucher et al., 2016), ή αιχμές (peaks) [Marcus et al., 2011; Scharl et al., 2012]. Στη μελέτη των Rohrdantz και συν. (2012) παρουσιάζεται μια τεχνική, η οποία υποστηρίζει την οπτική ανάλυση και αυτόματη εξαγωγή περιοχών δεδομένων, που παρουσιάζουν ιδιαίτερο ενδιαφέρον, συνδυάζοντας γραφικές αναπαραστάσεις ραβδογραμμάτων, με γραφήματα περιοχής. Το σύστημα TargetVue (Cao et al., 2016) χρησιμοποιεί τιμές πολικότητας συναισθήματος για την ανάλυση και απεικόνιση μη αναμενόμενης συμπεριφοράς χρηστών μέσω κοινωνικής δικτύωσης και ηλεκτρονικού ταχυδρομείου, αξιοποιώντας γραφικές αναπαραστάσεις γλύφου.
- 2) Η Ομαδοποίηση/ Κατηγοριοποίηση (Clustering/ Classification) περικλείει τεχνικές οι οποίες, πέραν της πραγματικής ανάλυσης συναισθημάτων, πραγματοποιούν κάποιου τύπου ομαδοποίησης ομαδοποίηση δεδομένων, π.χ. κάνοντας χρήση ετικετών (tagging),

και αναπαριστώντας τις ομαδοποιήσεις αυτές με γραφικό τρόπο. Στη μελέτη των Oelke και συν. (2009) παρουσιάζονται τεχνικές ανάλυσης σχολίων πελατών και οπτικοποίηση τους μέσω συνοπτικών αναφορών και κυκλικών χαρτών συσχετίσεων. Στο έργο των Brew και συν. (2011) παρουσιάζεται το σύστημα SentireCrowds, το οποίο αξιοποιεί τεχνικές ομαδοποίησης για τη γραφική απεικόνιση της χρονικής ανάλυσης πολικότητας συναισθημάτων χρηστών του twitter. Οι χρήστες ομαδοποιούνται σε δύο ομάδες με βάση σχόλια που έχουν κάνει και αφού πραγματοποιηθεί ο συγκεντρωτικός υπολογισμός πολικότητας για κάθε ομάδα ανά χρονικό βήμα. Για την οπτικοποίηση χρησιμοποιούνται γραφήματα περιοχής και σύνθετοι χάρτες-δέντρων (tree maps). Οι Kim και Lee (2014) παρουσίασαν μια τεχνική την οποία εφάρμοσαν σε σχόλια των πελατών. Η μέθοδος χρησιμοποιεί δισδιάστατες αναπαραστάσεις διασποράς με χρωματική κωδικοποίηση. Τέλος, στο σύστημα ToPIN (Sung et al., 2016) χρησιμοποιούνται αναπαραστάσεις κόμβων μεταβλητής φωτεινότητας για να οπτικοποιηθούν αποτελέσματα ανάλυσης σχολίων μαθητών.

- 3) Η Σύγκριση Πολλαπλών Οντοτήτων αποτελεί κατηγορία τεχνικών απεικόνισης, στην οποία εντάσσονται εκείνες οι περιπτώσεις που συγκρίνουν και αντιπαραβάλλουν με γραφικό τρόπο τα αποτελέσματα της ανάλυσης συναισθημάτων. Για παράδειγμα, στη μελέτη των Xu και συν. (2011) συγκρίνονται οι πολικότητες συγκεκριμένων χαρακτηριστικών προϊόντων, όπως αυτές προκύπτουν από τη σχετική ανάλυση συναισθημάτων, με χρήση γραφήματος κόμβων-διασυνδέσεων (node-link). Οι Kuksenok και συν. (2012) χρησιμοποιούν χρονική απεικόνιση γραμμής (timeline) για να οπτικοποιήσουν την εμφάνιση πολλαπλών κατηγοριών επιπτώσεων και να προσδιορίσουν τις μεταξύ τους συσχετίσεις. Στη μελέτη των Liu και συν. (2016) απεικονίζονται γραφικά σε «τροχούς μάρκας» (Brand Wheels) και συγκρίνοντας τα αποτελέσματα της ανάλυσης συναισθημάτων διαφόρων πτυχών μιας μάρκας (brand), αξιοποιώντας δεδομένα δημοσιεύσεων σε μέσα κοινωνικής δικτύωσης και σχόλια των εργαζομένων. Τέλος, απεικόνιση σύγκρισης χρησιμοποίησαν οι Kempter και συν. (2014) στο σύστημα EmotionWatch, το οποίο ανέπτυξαν για να συγκρίνουν τις συναισθηματικές αντιδράσεις μεταξύ δύο θεμάτων ή οντοτήτων, χρησιμοποιώντας δεδομένα του twitter, ενώ οι Chen και συν. (2017) πρότειναν το σύστημα Lariat, το οποίο οπτικοποιούσε τη σύγκριση

ερωτημάτων στο twitter, κάνοντας χρήση πολλαπλών χαρακτηριστικών, μεταξύ αυτών και της πολικότητας των δεδομένων.

- 4) Η Συγκεντρωτική Απεικόνιση (Overview) αναφέρεται σε τεχνικές οι οποίες επιτυγχάνουν να απεικονίζουν τη «μεγάλη εικόνα», οπτικοποιώντας μεγάλο αριθμό δεδομένων ή μέσω συγκεντρωτικών (aggregated) αναπαραστάσεων. Για παράδειγμα, το σύστημα City Sentiment (Wu et al., 2013) απεικονίζει ένα πλέγμα γλύφον φυσαλίδων, που αντιστοιχούν σε πόλεις της Κίνας, κωδικοποιώντας ταυτόχρονα στο ίδιο γράφημα τον αριθμό των δημοσιεύσεων/ σχολίων μέσω Weibo από κάθε πόλη αλλά και τη συνολική ανιχνευθείσα πολικότητά τους. Το Σύστημα SocialBrands (Liu et al., 2016) πραγματοποιεί συγκεντρωτική απεικόνιση των χαρακτηριστικών μιας μάρκας, όπως για παράδειγμα είναι ο «ενθουσιασμός» ή η «ειλικρίνεια». Αυτό επιτυγχάνεται με χρήση νέφους φυσαλίδων (bubble clouds). Άλλες τεχνικές που εμπίπτουν στη συγκεκριμένη κατηγορία παρουσιάζονται στις μελέτες των Miao και συν. (2009) και Scharl και συν. (2016), όπου χρησιμοποιούνται παραδοσιακά γραφήματα πίτας.
- 5) Η Συστηματική Παρακολούθηση (Monitoring) είναι τεχνική απεικόνισης η οποία στηρίζεται στη διαρκή δυναμική παρατήρηση και προειδοποίηση σε περίπτωση αλλαγών στα υπό ανάλυση δεδομένα. Κάποιες από τις εν λόγω τεχνικές χρησιμοποιούν μία κινούμενη γραμμή (timeline metaphor) ή άλλου είδους κινούμενη παράσταση (animation) ή συνδυασμό πολλαπλών γραφημάτων, που ενημερώνονται συνεχώς για να απεικονίσουν τις αλλαγές στα δεδομένα (Krstajic et al., 2012; Kranjc et al., 2015; Weiler et al., 2015; Gobron et al., 2010; Gali et al., 2012; El-Assady et al., 2016; Krcadinac et al., 2016; Marcus et al., 2011; Cui et al., 2012; Scharl et al., 2016; Calderon et al., 2014).
- 6) Η Πλοήγηση (Navigation) ή Εξερεύνηση (Exploration) ομαδοποιεί τεχνικές, που υποστηρίζουν διαδραστική πλοήγηση στα δεδομένα και εξερεύνησή τους. Χαρακτηριστικά παραδείγματα τεχνικών, που εμπίπτουν στη συγκεκριμένη κατηγορία, παρουσιάζονται στις μελέτες των Duan και συν. (2012), Brooks και συν. (2014), Li και συν. (2016), Lu και συν. (2016), Hao και συν. (2013), Torkildson και συν. (2014), Chen και συν. (2015), όπου χρησιμοποιούνται οπτικοποιήσεις δεδομένων γραμμής χρόνου (time line), χάρτη (map) ή άλλες αντίστοιχες απεικονίσεις.

7) Η Αντιμετώπιση Αβεβαιότητας (Uncertainty Tackling) εμπεριέχει τεχνικές απεικόνισης συναισθήματος, που εστιάζουν στη χρήση και απεικόνιση της αβεβαιότητας που υπάρχει στα υποκείμενα δεδομένα. Παραδείγματα τεχνικών αυτής της κατηγορίας παρουσιάζονται στις μελέτες των Annett και Kondrak (2008), Schäfer και Spurk (2010), Diakopoulos και συν. (2010) και Gamon και συν. (2008). Στο σύστημα OpinionSeer (Wu et al., 2010) μοντελοποιείται και οπτικοποιείται η αβεβαιότητα με χρήση συνδυασμού ακτινικών γραφημάτων (radial views) και γραφημάτων διασποράς (scatter plots), ώστε να αναπαρασταθούν ταυτόχρονα δεδομένα στη χρονική και γεωχωρικά διάσταση σε συνδυασμό με την πολικότητα, την αβεβαιότητα και το προφίλ των πελατών. Στη μελέτη των Rohrdantz και συν. (2012) επίσης πραγματοποιείται ανάλυση και οπτικοποίηση της αβεβαιότητας με χρήση γραφημάτων χρονικής πυκνότητας (time density plots), ενώ στη μελέτη των Makki και συν. (2014) η γραφική αναπαράσταση της αβεβαιότητας γίνεται μέσω νέφους δέντρων (tree cloud) και γραφημάτων διασποράς (scatter plots) με ενσωματωμένα νέφη λέξεων (word clouds). Η αβεβαιότητα σε αυτή την περίπτωση σχετίζεται με τις διαφορούμενες προβλέψεις πολικότητας, οι οποίες επισημαίνονται με κίτρινο χρώμα στην αναπαράσταση του νέφους δέντρων. Τέλος, στις μελέτες των Lu και συν. (2016) και Lu και συν. (2015) χρησιμοποιούνται αντίστοιχα για την απεικόνιση της αβεβαιότητας γραφήματα ροής (stream graphs) και χάρτης γλύφων (glyphs map).

2.5.5 Τεχνικές με βάση τα χαρακτηριστικά οπτικοποίησης

Η παρούσα ενότητα, εστιάζει στην οπτικοποίηση ανάλυσης συναισθημάτων με βάση τα χαρακτηριστικά οπτικοποίησης (visualization aspects) στις χρησιμοποιούμενες τεχνικές. Η πρώτη κατηγορία τεχνικών υπό αυτή την οπτική κατατάσσει τις τεχνικές με βάση τις ιδιότητες των γραφικών (visual variables), όπως παρουσιάζεται ακολούθως (Kucher et al., 2018):

1) Η οπτικοποίηση των συναισθημάτων χρησιμοποιώντας χρωματικές αναπαραστάσεις είναι ο δημοφιλέστερος τρόπος απεικόνισης και χρησιμοποιείται από πληθώρα τεχνικών. Μία πολύ συνήθης πρακτική είναι η απεικόνιση της πολικότητας συναισθημάτων με πράσινο χρώμα όταν πρόκειται για θετικά συναισθήματα και κόκκινο όταν αναφέρεται σε αρνητική πολικότητα (Wensel & Sood, 2008; Tsirakis et al., 2016). Σε κάποιες περιπτώσεις, χρησιμοποιείται αντίστροφος χρωματισμός και το πράσινο ή μπλε απεικονίζει αρνητικές

πολικότητες ενώ το κόκκινο ή πορτοκαλί θετικές (Wu et al., 2012; Wu et al., 2013; Calderon et al., 2014; Chen et al., 2014; Hoque & Carenini, 2014; Hoque & Carenini, 2016; Yu et al., 2016). Στην περίπτωση τεχνικών που χρησιμοποιούν κλίμακες πολικότητας πέραν των θετικών και αρνητικών, αυτό που συνήθως γίνεται είναι να χρησιμοποιείται μια προοδευτική χρωματική διαβάθμιση, ώστε να κωδικοποιηθούν όλες οι κατηγορίες ή διαστάσεις (Kang & Ren, 2011; Zhao et al., 2014a; El-Assady et al., 2016; Almutairi, 2013; Wang et al., 2015).

- 2) Η οπτικοποίηση με βάση τη Θέση ή τον Προσανατολισμό είναι επίσης μία πρακτική, η οποία συναντάται σε σημαντικό αριθμό τεχνικών. Ως παράδειγμα μπορεί να αναφερθεί μία απεικόνιση γραμμής (line plot), στην οποία ο κατακόρυφος άξονας απεικονίζει τη συνολική τιμή πολικότητας (Taskaya & Ahmad, 2003; Cervantes et al., 2015), ενώ σε άλλες περιπτώσεις χρησιμοποιούνται πολλαπλές γραμμές για την αναπαράσταση πολικότητας με την πάροδο του χρόνου (Hoeber et al., 2016). Εναλλακτικά, κάποιες φορές χρησιμοποιείται ο προσανατολισμός για να οπτικοποιηθεί η πολικότητα σε ένα ή σε δύο άξονες (Makki et al., 2014; Chen et al., 2015; Wensel & Sood, 2008; Zhao et al., 2016a).
- 3) Το μέγεθος ή η περιοχή ως έννοιες χρησιμοποιούνται επίσης για την οπτικοποίηση συναισθημάτων από μεγάλο αριθμό τεχνικών. Χαρακτηριστικά παραδείγματα αποτελούν οι τεχνικές που χρησιμοποιούν ραβδογράμματα (Alper et al., 2011; Mohammad et al., 2016), γραφήματα πίτας (Marcus et al., 2011; Das et al., 2012), διαγράμματα περιοχής ή ροής (Liu et al., 2013; Kranjc et al., 2015), ή γραφήματα αστεριών (Kempter et al., 2014; Munezero et al., 2015). Σε άλλες περιπτώσεις χρησιμοποιούνται κλιμακωτά επίπεδα λάμπης εντός μιας περιοχής (Gamon et al., 2008), ή κλιμακωτά μεγέθη γραμμάτων (Wecker et al., 2014).
- 4) Αντιπροσωπευτικές περιπτώσεις αυτής της κατηγορίας τεχνικών είναι οι αναπαραστάσεις γλύφου μόνες τους ή συνδυαστικά με άλλου τύπου γραφήματα (Liu et al., 2003; Gobron et al., 2010; Lee et al., 2010; Cresci et al., 2014). Οι Kim και Lee (2014) χρησιμοποιούν διάφορα σχήματα για επισήμανση, σε συνδυασμό με χρωματική κωδικοποίηση για την αναπαράσταση θετικών και αρνητικών κριτικών, ενώ σε άλλες περιπτώσεις έχουν χρησιμοποιηθεί σχήματα όπως ο χάρτης θερμότητας (heat map) [Li & Ren, 2009; Almutairi, 2013] και γραφήματα αστεριών (Kempter et al., 2014; Munezero et al., 2015).

5) Η Υφή ή το Μοτίβο είναι επίσης μεταβλητές οπτικοποίησης που χρησιμοποιούνται από κάποιες τεχνικές για την αναπαράσταση των συναισθημάτων. Για παράδειγμα στη μελέτη των Galí και συν. (2012) απεικονίζονται συγκεντρωτικά πολικότητες συναισθημάτων ανά μήνα ως υφασμένα μοτίβα (woven patterns), ενώ οι Zhang και συν. (2013) χρησιμοποιούν για την οπτικοποίηση συναισθημάτων από δεδομένα μέσω κοινωνικής δικτύωσης μια ειδική αναπαράσταση (electron cloud model), που στηρίζεται στην υφή γραμμών τροχιάς. Στη μελέτη των Kuang και συν. (2014) εισάγεται η χρήση ενός εργαλείου απεικόνισης, που οπτικοποιεί τις πολικότητες σε ένα θεματικό χάρτη (choropleth map), ενώ στη μελέτη των Krcadinac και συν. (2016) έχει προταθεί μια τεχνική απεικόνισης για τα συναισθήματα συνομιλιών (chats), η οποία προσομοιώνει κινούμενα σωματίδια με υφή και χρώματα που μεταβάλλονται κατά τη διάρκεια της συνομιλίας.

Η δεύτερη υποκατηγορία τεχνικών αυτού του τύπου, που παρουσιάζεται παρακάτω, είναι αυτές που ομαδοποιούνται με βάση τον τύπο των γραφικών (Kucher et al., 2018):

- 1) Μία από τις δημοφιλέστερες οπτικοποιήσεις που χρησιμοποιούνται, στηρίζεται σε τύπους γραφικών όπως τα γραφήματα Γραμμών (line plots) και Κινούμενων Ροών (rivers). Γραφήματα γραμμής έχουν χρησιμοποιηθεί κυρίως για την απεικόνιση δεδομένων που εξελίσσονται στο χρόνο όπως για παράδειγμα αυτά που παρουσιάζονται στις τεχνικές των Taskaya και Ahmad (2003), Mishne και De Rijke (2006), Kucher και συν. (2016), Scharl και συν. (2016), Fukuhara και συν. (2007) και Diakopoulos και συν. (2014). Γραφήματα κινούμενων ροών, τα οποία μοιάζουν με ποταμούς περιέχονται στα συστήματα VisTravel (Li et al., 2016), VISA (Duan et al., 2012), PEARL (Zhao et al., 2014a), OpinionFlow (Wu et al., 2014) και IdeaFlow (Wang et al., 2016). Άλλες τεχνικές, οι οποίες χρησιμοποιούν συνδυασμούς των γραφικών αυτής της κατηγορίας, έχουν παρουσιαστεί στις μελέτες των Subasic και Huettner (2001), Di Caro και Grella (2013) και Zhao και συν. (2016b).
- 2) Οπτικοποίηση συναισθημάτων με χρήση γραφημάτων Σημείων (Pixels), Περιοχής (Area) ή Πίνακα (Matrix). Κλασσική γραφική αναπαράσταση αυτής της κατηγορίας είναι τα γραφήματα πίτας (pie) που χρησιμοποιούνται στα Συστήματα AMAZING (Miao et al., 2009), TwitInfo (Marcus et al., 2011) και 5W (Das et al., 2012) για να απεικονίσουν τις διάφορες κατηγορίες πολικότητας. Στη μελέτη των Kumamoto και συν. (2014) χρησιμοποιούνται απλές απεικονίσεις, που εμπεριέχουν γραφήματα γραμμής και πίτας για

δεδομένα που αντλούνται από το twitter. Σε άλλες περιπτώσεις χρησιμοποιούνται γραφήματα ράβδων διαφόρων τύπων, όπως ιστογράμματα που χρησιμοποιούνται στα Συστήματα FAVe (Guzman et al., 2014), VisOHC (Kwon et al., 2016) και GeoSentiment (Pino et al., 2016), ή γραφήματα ράβδου-σωρού (stacked bar charts) που χρησιμοποιούνται στις μελέτες των Dehiya και Mueller (2016) και Mohammad και συν. (2016). Στη μελέτη των Wanner και συν. (2011) χρησιμοποιείται συνδυασμός γραφημάτων απεικόνισης ράβδου και σημείων σε συνδυασμό με χρωματική κωδικοποίηση της πολικότητας, ενώ στα συστήματα Pulse (Gamon et al., 2005) και SentireCrowds (Brew et al., 2011) χρησιμοποιείται γραφικό χάρτη δέντρων (tree map). Τέλος, οι Wu και συν. (2012) εισάγουν ένα καρτογράφημα (cartogram), που αποτελεί μέρος του συστήματος τους City Flow για να αναπαραστήσουν την πολικότητα των συναισθημάτων δεδομένων του Weibo που προέρχονται από συγκεκριμένες πόλεις.

- 3) Η οπτικοποίηση με χρήση γραφημάτων κόμβου-διασυνδέσεων (node-link) αποτελεί μία ακόμη συνήθη πρακτική σε τεχνικές απεικόνισης συναισθημάτων. Παραδείγματα τέτοιων γραφημάτων υπάρχουν στις δουλειές των Kang και Ren (2011), Small (2011) και Makki και συν. (2014). Επιπλέον, σε αρκετές περιπτώσεις οι συγκεκριμένες μορφές γραφικών συνδυάζονται με άλλες όπως χάρτες (Cao et al., 2012; Zhang et al., 2013), ή γραφήματα ράβδου-στοίβας (Xu et al., 2011; Hoque & Carenini, 2014; Hoque & Carenini, 2016). Άλλες ενδιαφέρουσες παραλλαγές τέτοιου τύπου απεικονίσεων είναι τα διαγράμματα τόξων (arc diagrams) και ο συνδυασμός τους με γραφήματα κινούμενων ροών (Chen et al., 2006; Fu et al., 2017).
- 4) Μία από τις πολύ ενδιαφέρουσες, σχετικά πρόσφατες κατηγορίες είναι αυτή που χρησιμοποιεί αναπαραστάσεις νέφους (cloud) ή γαλαξία (galaxy). Χαρακτηριστικό παράδειγμα τεχνικών αυτής της κατηγορίας είναι αυτές που χρησιμοποιούν τις διάφορες μορφές νέφους λέξεων (word clouds) ή ετικετών (tag clouds) [Yatani et al., 2011; Scharl et al., 2012]. Οι Wang, J. και συν. (2013) χρησιμοποιούν αναπαράσταση νέφους λέξεων για να απεικονίσουν την πολικότητα δεδομένων που προέρχονται από το twitter, λαμβάνοντας υπόψη και τη διάσταση του χρόνου. Σε άλλες περιπτώσεις χρησιμοποιούνται νέφη κουκκίδων ή αντίστοιχων συμβόλων επισήμανσης (Wang, C. et al., 2013; Kim & Lee, 2014). Στη μελέτη των Lu και συν. (2014) οι τιμές πολικότητας απεικονίζονται σε νέφη λέξεων και γραφήματα φυσαλίδων χρονικής εξέλιξης. Τέλος, στη μελέτη των Ha και

συν. (2014) παρουσιάζονται τα γραφήματα «γαλαξία». Επίσης, στη συγκεκριμένη κατηγορία χρησιμοποιούνται για την ανάλυση συναισθημάτων από κριτικές ταινιών, απεικονίζοντας ένα χάρτη αστερισμών των ταινιών σε συνδυασμό με αντιπροσωπευτικές της πολικότητας λέξεις που ανιχνεύονται στα σχόλια.

- 5) Η οπτικοποίηση συναισθημάτων με χρήση χαρτών (maps) είναι ένας επίσης αρκετά δημοφιλής τρόπος γραφικής αναπαράστασης της πολικότητας, ο οποίος μάλιστα συχνά είναι και διαδραστικός. Πρόκειται για τεχνικές που χρησιμοποιούν είτε πραγματικούς γεωγραφικούς χάρτες, είτε «λογικούς» χάρτες, που επιτρέπουν στους τελικούς χρήστες να εντοπίσουν και να κατανοήσουν περιοχές ενδιαφέροντος σε σχέση με τα αποτελέσματα των αναλύσεων. Επιπρόσθετα σε κάποιες περιπτώσεις, όπως το σύστημα TwitInfo των Marcus και συν. (2011), οι χάρτες συνοδεύονται από σύμβολα επισήμανσης, ή επικαλύπτονται με άλλα γραφικά για να ενσωματώσουν με παραστατικότερο τρόπο την απεικόνιση της πολικότητας (Zhang et al., 2013; Qiu et al., 2015; Dai & Prout, 2016). Το σύστημα Tweetviz, που παρουσιάστηκε από τους Sijtsma και συν. (2016), χρησιμοποιεί απεικόνιση χάρτη για να αναπαραστήσει τις περιοχές των επιχειρήσεων που αξιολογούνται μέσω δεδομένων του tweeter και κωδικοποιεί την πολικότητα χρωματικά πάνω στο χάρτη. Σε άλλες περιπτώσεις χρησιμοποιούνται θεματικοί χάρτες (choropleth). Παραδείγματα τέτοιων τεχνικών υπάρχουν στο σύστημα MoodLens (Zhao et al., 2012), στο σύστημα ImgWordle (Kuang et al., 2014), αλλά και στη δουλειά των Yu και συν. (2016). Στη μελέτη των Zhao και συν. (2016b) παρουσιάζεται το σύστημα Social Sentiment Sensor, το οποίο χρησιμοποιεί πολλαπλές γραφικές αναπαραστάσεις για την απεικόνιση των αποτελεσμάτων ανάλυσης συναισθημάτων δεδομένων ιστολογίων, μεταξύ άλλων γραφήματα γραμμής, πίτας και ράβδου, θεματικούς χάρτες και γλύφους. Στην κατηγορία των «λογικών» χαρτών ανήκουν καταρχάς οι χάρτες θερμότητας (heat maps). Παραδείγματα τέτοιων τεχνικών περιλαμβάνονται στις μελέτες των Hennig και συν. (2014) και Hennig και συν. (2015). Σε κάποιες περιπτώσεις χρησιμοποιούνται συνδυασμοί heat maps με άλλες αναπαραστάσεις (Hao et al., 2013; Steed et al., 2015; Pino et al., 2016).
- 6) Όσο παράδοξο κι αν ακούγεται μετά από την παράθεση όλων των ανωτέρω τεχνικών, σε αρκετές περιπτώσεις χρησιμοποιείται κείμενο (text) ως μέσο οπτικοποίησης των αποτελεσμάτων ανάλυσης συναισθημάτων. Παραδείγματα είναι η τεχνική Ink Blots από τους Abbasi και Chen (2007), τα συστήματα Jigsaw (Gorg et al, 2013), Semantize (Wecker

et al., 2014) και uVSAT (Kucher et al., 2016). Άλλες τεχνικές που στηρίζονται στο κείμενο για οπτικοποίηση παρουσιάζονται στις έρευνες των Chen και συν. (2015 και Brath και Banissi (2016).

- 7) Μία ακόμη κατηγορία τεχνικών οπτικοποίησης δομείται από εκείνες τις τεχνικές που χρησιμοποιούν γλύφους ή εικονίδια για την παρουσίαση των αποτελεσμάτων ανάλυσης συναισθημάτων. Όπως έχει ήδη αναφερθεί, τα γραφήματα γλύφων, μόνα τους ή σε συνδυασμό με άλλες οπτικές αναπαραστάσεις, χρησιμοποιούνται σε ευρεία έκταση. Παραδείγματα τέτοιων οπτικοποιήσεων παρουσιάζονται στις μελέτες των Wanner και συν. (2009) και Wanner και συν. (2012). Σε κάποιες περιπτώσεις υψηλής πυκνότητας γλύφων, το τελικό οπτικό αποτέλεσμα είναι παρόμοιο με αναπαραστάσεις που βασίζονται σε pixel. Για παράδειγμα στο σύστημα FluxFlow (Zhao et al., 2014a) τα αποτελέσματα της ανάλυσης απεικονίζονται ως γλύφοι νημάτων (thread glyphs). Άλλες τεχνικές χρησιμοποιούν γλύφους-μετρητές (Wensel & Sood, 2008; Zhao et al., 2016b), ή άλλες παραλλαγές (Liu et al., 2003; Cresci et al., 2014; Gobron et al., 2010).

2.6 Εργαλεία και λογισμικό οπτικοποίησης ανάλυσης συναισθημάτων

Στην παρούσα ενότητα επιχειρείται μία καταγραφή εργαλείων (εμπορικών και ανοιχτού κώδικα – open source), τα οποία δύναται να χρησιμοποιηθούν για την οπτικοποίηση αποτελεσμάτων ανάλυσης συναισθημάτων. Τα εργαλεία που αναφέρονται στον Πίνακα 2-2 είναι μία σύνθεση λογισμικού και εφαρμογών, που είναι κατάλληλα για γραφικές αναπαραστάσεις δεδομένων (π.χ. όλα τα εργαλεία και το λογισμικό επιχειρησιακής ευφυΐας – Business Intelligence, BI), εργαλεία text mining, επεξεργασίας φυσικής γλώσσας (Natural Language Processing, NLP) και εξειδικευμένης οπτικοποίησης ανάλυσης συναισθημάτων. Τα συγκεκριμένα εργαλεία χρησιμοποιούν πλήθος διαφορετικών τεχνικών ανάλυσης, μεταξύ των οποίων αλγόριθμους στατιστικής ανάλυσης και ομαδοποίησης (statistical analysis and clustering), τεχνητής νοημοσύνης (artificial intelligence) και μηχανικής και σε βάθος μάθησης (machine and deep learning). Επιπλέον, υφίσταται επίσης μεγάλο πλήθος εφαρμογών που έχουν αναπτυχθεί πάνω σε ανοιχτού κώδικα προγραμματιστικές πλατφόρμες και εργαλεία (π.χ. Python, R, κ.λπ.), οι οποίες επίσης δεν περιλαμβάνονται, μολονότι μία τέτοια υλοποίηση έχει πραγματοποιηθεί και προτείνεται στο πλαίσιο της παρούσας εργασίας.

1.	Apache OpenNLP	Λογισμικό ανοιχτού κώδικα με προηγμένες δυνατότητες ανάλυσης κειμένου και λειτουργικότητες οπτικοποίησης των αποτελεσμάτων.
2.	Aylien	Χρησιμοποιεί τεχνολογίες τεχνητής νοημοσύνης, μηχανικής μάθησης και επεξεργασίας φυσικής γλώσσας για την ανάλυση κειμένου. Εστιάζει σε περιοχές στρατηγικών marketing, υποστήριξης πελατών και ανάπτυξης προϊόντων και παρέχει λειτουργικότητες αυτοματοποιημένων αναλύσεων συναισθημάτων και γραφικής απεικόνισής τους. Επίσης, περιλαμβάνει API ειδήσεων.
3.	Bismart Folksonomy	Πρόκειται για λογισμικό επεξεργασίας κειμένου, που μπορεί να διαχειριστεί αποτελεσματικά αδόμητες πληροφορίες και να κάνει ανάλυση και οπτικοποίηση με ταχύτητα και φιλικό προς το χρήστη τρόπο.
4.	GATE - Java NLP tools	Σουίτα με δυνατότητες επεξεργασίας γλώσσας και ανάλυσης κειμένου σε διάφορες γλώσσες. Μεταξύ άλλων παρέχει διαφορετικές μορφές οπτικοποίησης.
5.	Google Cloud NLP/ Google Fusion Tables	Εργαλεία ανάλυσης μεγάλων όγκων κειμένου και οπτικοποίησης μεγάλων όγκων δεδομένων και αποτελεσμάτων ανάλυσης συναισθημάτων, που υποστηρίζονται από την πλατφόρμα της Google. Χρησιμοποιούνται μεταξύ άλλων και τεχνικές μηχανικής μάθησης.
6.	IBM Watson	Δημοφιλής, εμπορικά διαθέσιμη σουίτα τεχνητής νοημοσύνης με προηγμένες δυνατότητες ανάλυσης και οπτικοποίησης δομημένων και αδόμητων δεδομένων. Μεταξύ άλλων εμπεριέχει λειτουργικότητες επεξεργασίας φυσικής γλώσσας, ανάλυσης και κατηγοριοποίησης συναισθημάτων, μοντέλα μηχανικής και «σε βάθος» μάθησης και μοντέρνες διαδραστικές οπτικοποιήσεις.
7.	Infogram	Παρέχει τη δυνατότητα απεικόνισης δεδομένων σε πολλά διαφορετικά είδη διαδραστικών γραφημάτων.
8.	MeaningCloud	Λύση που εμπεριέχει πολλαπλές αυτοματοποιημένες δυνατότητες ανάλυσης συναισθημάτων και γραφικής απεικόνισής τους. Ολοκληρώνεται με εργαλεία όπως τα RapidMiner, Zapier, Google Sheets, και Excel, μέσω εξειδικευμένων APIs.
9.	Microsoft Cognitive Services	Η σουίτα γνωσιακών υπηρεσιών της Microsoft περιλαμβάνει αλγόριθμους τεχνητής νοημοσύνης και Application Programming Interfaces (APIs) που μπορούν να χρησιμοποιηθούν για την ανάπτυξη εφαρμογών ανάλυσης και οπτικοποίησης κειμένου.
10.	MonkeyLearn	Εύχρηστη πλατφόρμα ανάλυσης συναισθημάτων, που εμπεριέχει έτοιμα εκπαιδευμένα text analytics μοντέλα. Επιπρόσθετα, περιλαμβάνει σουίτα οπτικοποίησης δεδομένων με έτοιμα πρότυπα και ποικίλες δυνατότητες γραφικής απεικόνισης. Τέλος περιλαμβάνονται αλγόριθμους μηχανικής μάθησης και δυνατότητες ολοκλήρωσης (integration) με πολλά δημοφιλή εργαλεία και λογισμικά (π.χ. Excel, Google Sheets, RapidMiner, κ.λπ.).
11.	Open text	Λογισμικό ανάλυσης συναισθημάτων και οπτικοποίησης των αποτελεσμάτων μέσω διαδραστικών γραφημάτων.
12.	Opinion Crawl	Online εργαλείο ανάλυσης συναισθημάτων, που χρησιμοποιεί πολλαπλές πηγές όπως προϊόντα, εταιρίες, ανθρώπους και συμβάντα. Εμπεριέχει δυνατότητες οπτικοποίησης των αναλύσεων μέσω γραφημάτων και μάλιστα σε πραγματικό χρόνο.
13.	Orange	Πλατφόρμα ανοιχτού κώδικα με διαδραστικές ροές, που υποστηρίζει την ανάλυση κειμένων και οπτικοποίηση των αποτελεσμάτων, χρησιμοποιώντας και τεχνικές μηχανικής μάθησης.
14.	QDA Miner	Εύχρηστο λογισμικό που μπορεί να χρησιμοποιηθεί για την ανάλυση κειμένου ειδήσεων, συνεντεύξεων, απαντήσεων καταναλωτών και άλλων. Υποστηρίζει πολλαπλές μορφές κειμένου (π.χ. pdf, rtf, html, csv, κ.λπ.) και παρέχει πολλαπλές δυνατότητες οπτικοποίησης και εξαγωγής των δεδομένων και των αποτελεσμάτων της ανάλυσης.

15.	Qlik	Δημοφιλές εργαλείο για τη δημιουργία και απεικόνιση διαδραστικών γραφημάτων, dashboards, και σχετικών εφαρμογών, μεταξύ των οποίων και εφαρμογών ανάλυσης συναισθημάτων.
16.	Quick Search	Πρόκειται για μηχανή αναζήτησης κοινωνικών δικτύων, η οποία αναλύει τεράστιους όγκους διαλόγων και παρέχει αποτελέσματα όπως KPIs μιας μάρκας (brand), τάσεις σε πραγματικό χρόνο, συγκρίσεις, κ.λπ.
17.	RapidMiner	Λογισμικό ανοιχτού κώδικα με λειτουργικότητες ανάλυσης και οπτικοποίησης κειμένου. Αντλεί δεδομένα από πολλαπλές πηγές όπως online αξιολογήσεις, δημοσιεύματα και μηνύματα μέσω κοινωνικής δικτύωσης.
18.	Semantria/ Lexanalytics	Εξειδικευμένες εφαρμογές ανάλυσης και οπτικοποίησης συναισθημάτων με δυνατότητα απευθείας επεξεργασίας κειμένου και δημοσιεύσεων στα μέσα κοινωνικής δικτύωσης (Lexanalytics). Οι εν λόγω υπηρεσίες προσφέρονται και μέσω APIs και σχετικού Excel plugin και χρήσης deep learning (Semantria).
19.	SAS Sentiment Analysis	Ευρέως διαδεδομένο εξειδικευμένο εμπορικό λογισμικό, που πραγματοποιεί με αυτοματοποιημένο τρόπο ανάλυση και οπτικοποίηση συναισθημάτων σε πραγματικό χρόνο, χρησιμοποιώντας συνδυασμό στατιστικής μοντελοποίησης και επεξεργασίας φυσικής γλώσσας με χρήση κανόνων.
20.	Salience	Βιβλιοθήκες ανάλυσης κειμένου, που μπορούν να ολοκληρωθούν με εφαρμογές χρηστών. Παρέχονται σημαντικές δυνατότητες όπως ανάλυση συναισθημάτων, ανάλυση προθέσεων, σήμανση μέρους του λόγου (part-of-speech tagging) και αναγνώριση γλώσσας.
21.	Tableau	Δημοφιλές εργαλείο BI που δίνει τη δυνατότητα προηγμένων γραφικών αναπαραστάσεων δεδομένων μεταξύ των οποίων και αποτελεσμάτων ανάλυσης συναισθημάτων.
22.	Texminer	Εργαλείο ανάλυσης κειμένου που δουλεύει με αρχεία τύπου txt και pdf. Υποστηρίζει πολλαπλές γλώσσες και έχει δυνατότητες οπτικοποίησης ανάλυσης συναισθημάτων.
23.	Thematic	Σουίτα εργαλείων που χρησιμοποιούν τεχνολογίες τεχνητής νοημοσύνης, επεξεργασίας φυσικής γλώσσας, μηχανικής και «σε βάθος» μάθησης για την ανάλυση κειμένου. Δίνει τη δυνατότητα για πολλαπλές οπτικοποιήσεις ενδεικτικές μεταξύ άλλων για την ανάλυση συναισθημάτων. Ολοκληρώνεται με εργαλεία όπως τα Zendesk, SurveyMonkey.
24.	Trackur	Λογισμικό που υποστηρίζει αυτοματοποιημένη ανάλυση επιρροής και συναισθημάτων και οπτικοποίηση αυτών διατρέχοντας κείμενα, μέσα κοινωνικής δικτύωσης και δημοσιευμένα νέα.
25.	Voyant Tools	Πρόκειται για εργαλεία ανάλυσης και οπτικοποίησης κειμένου από το διαδίκτυο. Δεν έχουν το βάθος και τις δυνατότητες άλλων αντίστοιχων εργαλείων, αλλά παρέχουν γρήγορες και περιεκτικές οπτικοποιήσεις συναισθημάτων.

Πίνακας 2-2: Εργαλεία και λογισμικό οπτικοποίησης ανάλυσης συναισθημάτων.

3 Μελέτη Περίπτωσης: Ανάπτυξη Διαδραστικής Διεπαφής Παρουσίασης Αποτελεσμάτων Ανάλυσης Συναισθημάτων

3.1 Εισαγωγή

Στην παρούσα ενότητα παρουσιάζεται ως μελέτη περίπτωσης η ανάπτυξη μιας διαδραστικής διεπαφής, μέσω της οποίας παρουσιάζονται, αναλύονται και αξιολογούνται τα αποτελέσματα ανάλυσης συναισθημάτων. Για το σκοπό αυτό έχει χρησιμοποιηθεί ένα σύνολο δεδομένων (data set), το οποίο περιλαμβάνει κριτικές και αξιολογήσεις πελατών για προϊόντα και υπηρεσίες που παρέχονται από μια εταιρία. Για την ανάλυση συναισθημάτων και οπτικοποίηση των αποτελεσμάτων έχει χρησιμοποιηθεί η γλώσσα Python, περιλαμβανομένων και δημοφιλών βιβλιοθηκών γραφικής απεικόνισης όπως είναι η Plotly και η Dash. Οι τύποι των γραφημάτων που δημιουργούνται είναι ενδεδειγμένοι για να οπτικοποιήσουν με αποτελεσματικό τρόπο τα αποτελέσματα των αναλύσεων.

3.2 Αντικείμενο υλοποίησης

Η οπτική αναπαράσταση των συναισθημάτων, που προκύπτουν από την ανάλυση του περιεχομένου εγγράφων κειμένου με πολλούς διαφορετικούς τρόπους, θα αποτελέσει το αντικείμενο υλοποίησης της παρούσας εργασίας. Το κείμενο περιλαμβάνει κριτικές και αξιολογήσεις των προϊόντων και παρεχόμενων υπηρεσιών μιας επιχείρησης. Στην υλοποίηση που έχει πραγματοποιηθεί εξετάζονται με οπτικό τρόπο τα συναισθήματα που εμπεριέχονται στο περιεχόμενο των εγγράφων από διαφορετικές πτυχές και σε διαφορετικά επίπεδα λεπτομέρειας, επισημαίνονται συγκεκριμένες αντιπροσωπευτικές λέξεις και θέματα, και ανιχνεύονται συγκεκριμένα συναισθήματα στα γεγονότα που παρουσιάζονται.

Η μετατροπή μη δομημένων (unstructured) δεδομένων κειμένου σε δομημένη (structured) προς απεικόνιση πληροφορία είναι μία από τις βασικότερες προκλήσεις που πρέπει να αντιμετωπιστεί σε περιπτώσεις όπως η οπτικοποίηση συναισθημάτων. Στη μελέτη περίπτωσης που υλοποιείται στην παρούσα εργασία χρησιμοποιείται για το σκοπό αυτό η γλώσσα Python, αξιοποιώντας για τη γραφική απεικόνιση της ανάλυσης συναισθημάτων συγκεκριμένες

βιβλιοθήκες γραφικών όπως είναι η Plotly και η Dash. Πολλές από τις απεικονίσεις κειμένου που πραγματοποιούνται δεν αντιπροσωπεύουν απευθείας το κείμενο αλλά τα αποτελέσματα μοντέλων ανάλυσης συναισθημάτων. Οι τεχνικές που έχουν χρησιμοποιηθεί για την οπτικοποίηση των συναισθημάτων αξιοποιούν λειτουργίες όπως είναι η καταμέτρηση λέξεων, το μήκος των προτάσεων, οι ακολουθίες λέξεων που χρησιμοποιούνται, κ.λπ.

Η συγκεκριμένη μελέτη περίπτωσης περιλαμβάνει (1) την προετοιμασία, ανάλυση και παρουσίαση ενός συνόλου δεδομένων κειμένου που περιλαμβάνουν κριτικές και αξιολογήσεις των προϊόντων και παρεχόμενων υπηρεσιών μιας επιχείρησης από πελάτες της, και (2) την παρουσίαση των αποτελεσμάτων ανάλυσης συναισθημάτων, όπως είναι η πολικότητα μεμονωμένα ή σε συνδυασμό με άλλες πτυχές και χαρακτηριστικά. Η απεικόνιση είναι διαδραστική και περιλαμβάνει:

- 1) Τη μονοδιάστατη (single-variable ή univariate) απεικόνιση κατά την οποία ουσιαστικά οπτικοποιείται μία μόνο μεταβλητή. Πρόκειται για έναν από τους απλούστερους τύπους απεικόνισης που αποτελείται από παρατηρήσεις σε ένα μόνο χαρακτηριστικό ή πτυχή. Οι τύποι γραφημάτων που χρησιμοποιούνται στην περίπτωση αυτή, περιλαμβάνουν γραφήματα ράβδων (bar charts). Παραδείγματα τέτοιων απεικονίσεων είναι η κατανομή της πολικότητας συναισθημάτων, καθώς και η κατανομή άλλων χαρακτηριστικών όπως το score αξιολόγησης, το μήκος κειμένου, ο αριθμός των λέξεων, κ.λπ. Για κάποιες κατηγορίες χαρακτηριστικών χρησιμοποιούνται απλά γραφήματα ράβδων προς παρουσίαση για παράδειγμα της συχνότητας εμφάνισης. Σε κάποιες άλλες περιπτώσεις απεικόνισης χαρακτηριστικών κειμένου, η οπτικοποίηση στηρίζεται σε σύνολα n λέξεων ($n \geq 1$), γνωστά και ως n -grams που χρησιμοποιούνται ως σημεία παρατήρησης. Για παράδειγμα, τα μονογράμματα (unigrams) χρησιμοποιούν μία λέξη, τα διπλογράμματα (bigrams) χρησιμοποιούν δύο λέξεις, και τα τριπλογράμματα (trigrams) τρεις λέξεις, κ.λπ. Για τις προαναφερθείσες περιπτώσεις παρουσιάζονται και συγκρίνονται οπτικοποιήσεις επιμέρους χαρακτηριστικών. Μία άλλη τεχνική οπτικοποίησης που έχει υλοποιηθεί είναι η απεικόνιση ετικετών, που κωδικοποιούν συγκεκριμένα τμήματα του κειμένου (Part of Speech Tagging, POS). Η συγκεκριμένη τεχνική υλοποιείται με τη χρήση ενός απλού TextBlob API και χρησιμοποιεί ένα box plot για την απεικόνιση και σύγκριση της πολικότητας συναισθημάτων, ή άλλων επιμέρους χαρακτηριστικών του κειμένου.

- 2) Η ταυτόχρονη απεικόνιση δύο χαρακτηριστικών (bivariate visualization) οπτικοποιεί και συσχετίζει δύο επιμέρους πτυχές της ανάλυσης συναισθημάτων. Χαρακτηριστικό παράδειγμα γραφήματος, που χρησιμοποιείται για αυτή την κατηγορία, είναι το γράφημα πυκνότητας δύο διαστάσεων (2D density joint plot), που απεικονίζει ταυτόχρονα δύο χαρακτηριστικά. Στην παρούσα υλοποίηση περιλαμβάνεται η δημιουργία τέτοιων γραφημάτων, όπως για παράδειγμα είναι η κατανομή της πολικότητας σε συνάρτηση με τις κριτικές πελατών.
- 3) Η απεικόνιση και ανάλυση συσχετίσεων επιμέρους χαρακτηριστικών. Για το σκοπό αυτό στην παρούσα υλοποίηση χρησιμοποιούνται οι απαιτούμενες βιβλιοθήκες της Python Scattertext και spaCy. Παράδειγμα τέτοιου γραφήματος που παρουσιάζεται είναι αυτό που απεικονίζει το βαθμό συσχέτισης επιμέρους δραστηριοτήτων μιας εταιρίας με τις αξιολογήσεις πελατών.

3.3 Αναλυτικός σχεδιασμός διεπαφής – Δομή και βασικά Χαρακτηριστικά

Στην ενότητα αυτή παρουσιάζονται τα βασικά χαρακτηριστικά και η δομή της διαδραστικής διεπαφής (interface), που έχει αναπτυχθεί στο πλαίσιο της παρούσας μελέτης περίπτωσης. Για τη συγκεκριμένη διεπαφή ισχύουν τα ακόλουθα:

- 1) Έχει χρησιμοποιηθεί ένα σύνολο δεδομένων (data set), το οποίο αποτελείται από 20.161 εγγραφές (records) και περιλαμβάνει κριτικές και αξιολογήσεις πελατών για προϊόντα και υπηρεσίες, που παρέχονται από εταιρία πωλήσεων γυναικείων ρούχων και αξεσουάρ.
- 2) Η διεπαφή είναι προσπελάσιμη και απεικονίζει τα γραφικά μέσω περιηγητή ιστοσελίδων (web browser).
- 3) Για το κείμενο που αναλύεται, απεικονίζονται πολλαπλά γραφικά και πληροφορίες, τα οποία συνδυάζονται και ομαδοποιούνται με λογικό τρόπο ανά σελίδα.
- 4) Η εφαρμογή είναι πολλαπλών σελίδων (multi page application), οι οποίες έχουν μια λογική αλληλουχία.
- 5) Οι λογικές ενότητες/ σελίδες της web-εφαρμογής, που ομαδοποιούν τις πληροφορίες και τα γραφικά τα οποία απεικονίζονται μαζί, είναι οι ακόλουθες:

- a) Ενότητα A – Data Sample Basics: Εισαγωγική σελίδα, η οποία περιλαμβάνει:
- i. βασικές πληροφορίες για το σύνολο δεδομένων (data set) και σχετικό δείγμα
 - ii. σχετικό link προς την πηγή του αρχείου, όπου εμπεριέχονται αναλυτικές πληροφορίες για τα συγκεκριμένα δεδομένα.
 - iii. Βασικά συγκεντρωτικά στατιστικά του συνόλου δεδομένων, ώστε να παρέχεται μία συνολική αρχική εικόνα των δεδομένων. Ειδικότερα, τον Αριθμό Κριτικών (Number of Reviews), τη Μέση Ηλικία Πελατη (Mean of Customer Age), την Ελάχιστη Ηλικία Πελατη (Minimum Customer Age), τη Μέγιστη Ηλικία Πελατη (Maximum Customer Age), την Τυπική Απόκλιση της Ηλικίας Πελάτη (Standard Deviation of Customer Age), το Μέσο Μήκος Κριτικής (Average Review Length), το Ελάχιστο Μήκος Κριτικής (Minimum Review Length), το Μέγιστο Μήκος Κριτικής (Maximum Review Length), οι Βαθμοί Αξιολόγησης (Ratings), τα Μεγέθη Ρούχων (Sizes), και οι Τύποι Ρούχων (Item Types).
- b) Ενότητα B - Visualizations: Ομάδα ραβδογραμμάτων (bar charts), που οπτικοποιούν τη στατιστική ανάλυση των κριτικών και την κατανομή τους βάσει διαφορετικών κριτηρίων. Ειδικότερα, τα επιμέρους γραφικά αυτής της ενότητας είναι τα εξής:
- i. Ομάδα I: Κατανομή κριτικών βάσει διαφορετικών κριτηρίων, τα οποία επιλέγονται μέσω ξεχωριστών καρτελών (tabs). Ειδικότερα, έχουν υλοποιηθεί και εμφανίζονται τα κάτωθι γραφήματα:
 - Κατανομή των κριτικών ανά βαθμολογία (Review Rating).
 - Κατανομή των κριτικών βάσει πολικότητας (Polarity).
 - Κατανομή των κριτικών ανά ηλικία των κριτικών (Reviewer Age).
 - Κατανομή των κριτικών βάσει του μήκους κειμένου (Text Length).
 - Κατανομή των κριτικών βάσει αριθμού λέξεων (Word Count).
 - Κατανομή των κριτικών βάσει του δείκτη σύστασης (Recommended Indicator).
 - ii. Ομάδα II: Απεικόνιση λέξεων με μεγαλύτερη συχνότητα εμφάνισης βάσει τεσσάρων (4) συνδυαστικών κριτηρίων, τα οποία είναι το Μέγεθος (Size), τον Δείκτη Σύστασης (Recommended Indicator), το Σκορ Κριτικής (Purchase Rating)

και ο Τύπος Ρούχου (Item Type). Όλα τα προαναφερθέντα φίλτρα επιλέγονται μέσα από λίστες τιμών (list boxes). Ειδικότερα, έχουν υλοποιηθεί και εμφανίζονται τα κάτωθι γραφήματα:

- Οι 10 κορυφαίες λέξεις κατόπιν εξαίρεσης των κοινών λέξεων (stop words).
- Οι 10 κορυφαίες φράσεις τριών λέξεων (trigrams) κατόπιν εξαίρεσης των κοινών λέξεων (stop words).
- Τα 10 κορυφαία τμήματα επισήμανσης λόγου (speech tagging).

Τα γραφήματα ενημερώνονται δυναμικά με βάση τις εκάστοτε επιλογές φίλτρων από το χρήστη της εφαρμογής.

c) Ενότητα Γ – Box Plots: Γραφήματα τύπου box plot, που απεικονίζουν με ιδιαίτερα παραστατικό και διαδραστικό τρόπο τα συναισθήματα για τους επιμέρους τύπους ρούχων ή σκορ κριτικής, βάσει διαφορετικών κριτηρίων. Ειδικότερα, έχουν υλοποιηθεί και δύο Ομάδες γραφημάτων:

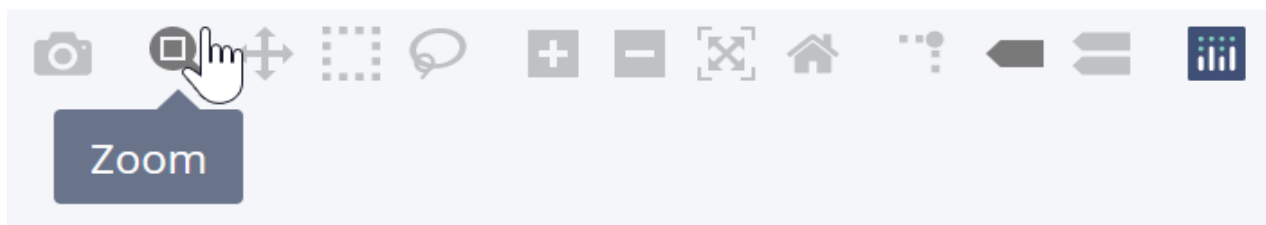
- i. *Ομάδα I*: Περιλαμβάνει δύο γραφήματα, τα οποία επιλέγονται μέσω ξεχωριστών καρτελών (tabs):
 - Κατανομή Πολικότητας (Polarity) ανά Τύπο Ρούχου (Item Type) για όλους τους διαθέσιμους τύπους (top, dress, accessories, bottom, jacket, trendy).
 - Κατανομή Μήκους Κριτικής (Review Length) ανά Τύπο Ρούχου (Item Type) για όλους τους διαθέσιμους τύπους (top, dress, accessories, bottom, jacket, trendy).
- ii. *Ομάδα II*: Περιλαμβάνει δύο γραφήματα, τα οποία επιλέγονται μέσω ξεχωριστών καρτελών (tabs):
 - Κατανομή Πολικότητας (Polarity) ανά Σκορ Κριτικής (Purchase Rating) για όλα τα διαθέσιμα σκορ (1 έως 5).
 - Κατανομή Ηλικίας Πελάτη (Customer Age) ανά Σκορ Κριτικής (Purchase Rating) για όλα τα διαθέσιμα σκορ (1 έως 5).

d) Ενότητα Δ – Latent Analysis: Γραφική απεικόνιση λανθάνουσας ανάλυσης (latent analysis), που παρουσιάζει τα έξι (6) δημοφιλέστερα θέματα συζήτησης (topics) μέσα από δύο διαφορετικούς τύπους γραφημάτων:

- 1) Γράφημα ράβδων
- 2) Γλύφος με t-SNE ομαδοποίηση (clustering).

Η διεπαφή έχει σχεδιαστεί με τέτοιο τρόπο ώστε να είναι διαδραστική. Η διαδραστικότητα διασφαλίζεται από τις ακόλουθες λειτουργίες:

- 1) Παρέχεται δυνατότητα πλοήγησης σε browser, από σελίδα σε σελίδα, μέσα στην εφαρμογή κατ' επιλογή χρήστη.
- 2) Τα γραφήματα είναι στη συντριπτική τους πλειοψηφία δυναμικά ανάλογα με τις επιλογές χρήστη, που θα πραγματοποιούνται είτε μέσω των σχετικών καρτελών (tabs) είτε από λίστες τιμών (list boxes).
- 3) Με τη μετακίνηση του ποντικιού πάνω στα διάφορα σημεία περιεχομένου των γραφημάτων απεικονίζονται πληροφορίες, που επεξηγούν τα απεικονιζόμενα δεδομένα, ή ενδιαφέροντα στατιστικά στοιχεία.
- 4) Παρέχεται δυνατότητα διάδρασης με το χρήστη μέσα από τα controls, που εμφανίζονται από την plotly, όπως αυτά απεικονίζονται παρακάτω.



Εικόνα 3-1: Λειτουργίες διάδρασης της διεπαφής.

Ειδικότερα, τα controls της διάδρασης είναι τα ακόλουθα:

1. «Κατέβασμα» γραφήματος ως εικόνα (download plot as a png)
2. Εστίαση (zoom)
3. Κεντράρισμα γραφήματος σε άλλο σημείο (Pan)

4. Επιλογή πλαισίου (box select)
5. Επιλογή λάσου (lasso select)
6. Μεγέθυνση (zoom in)
7. Σμίκρυνση (zoom out)
8. Επαναφορά μεγέθους στο αρχικό (auto scale)
9. Επαναφορά αξόνων στην αρχική μορφή (reset axis)
10. Εναλλαγή γραμμών ακραίων σημείων (toggle spike lines)
11. Εμφάνιση κοντινότερων δεδομένων (show closest data on hover)
12. Σύγκριση δεδομένων (compare data on hover)

3.4 Περιγραφή συνόλου δεδομένων

Στο πλαίσιο της συγκεκριμένης μελέτης περίπτωσης έχει χρησιμοποιηθεί σύνολο δεδομένων (Data Set) από το χώρο του ηλεκτρονικού εμπορίου και ειδικότερα από ηλεκτρονικό κατάστημα πώλησης γυναικείου ρουχισμού και αξεσουάρ. Το πηγαίο αρχείο διατίθεται από την Kaggle και είναι προσβάσιμο από το σύνδεσμο <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>. Έχει τον τίτλο «*Women's E-Commerce Clothing Reviews*», είναι μορφής .csv και περιέχει περισσότερες από 23.000 εγγραφές, που είναι κριτικές πελατών που έχουν πραγματοποιήσει αγορές. Ειδικότερα, περιέχει έντεκα (11) στήλες με επιμέρους πληροφορίες σχετικά με τις κριτικές που έχουν γίνει.

Για την εφαρμογή που έχει αναπτυχθεί στο πλαίσιο της παρούσας εργασίας, το αρχικό σύνολο δεδομένων έχει περάσει από ένα απαραίτητο στάδιο καθαρισμού και προεπεξεργασίας των δεδομένων, ώστε να καταστεί ευκολότερη και εγκυρότερη η όποια απεικόνιση πραγματοποιείται στη συνέχεια. Μέσα από αυτή τη διαδικασία έχουν απορριφθεί όσες από τις εγγραφές δεν ακολουθούσαν την προδιαγεγραμμένη γραμμογράφιση. Μετά την επεξεργασία αυτή, το τελικό αρχείο που χρησιμοποιείται περιλαμβάνει 20.161 εγγραφές, όπως άλλωστε προκύπτει και από την ανάλυση των σχετικών αποτελεσμάτων που παρουσιάζεται στην ενότητα 3.7. Αναλυτικά, οι πληροφορίες που περιέχονται στο αρχείο παρουσιάζονται στον Πίνακα 3-1.

	Όνομα Πεδίου	Περιγραφή
1.	Review ID	Μοναδικός κωδικός για κάθε αξιολόγηση
2.	Item ID	Κωδικός προϊόντος για το οποίο γίνεται η αξιολόγηση
3.	Customer Age	Ηλικία πελάτη
4.	Review Title	Τίτλος αξιολόγησης
5.	Review Text	Κείμενο αξιολόγησης
6.	Purchase Rating	Σκορ αξιολόγησης
7.	Recommended Indicator	Δείκτης σύστασης
8.	Number of Positive Feedbacks	Αριθμός θετικών αναδράσεων
9.	Size	Μέγεθος
10.	Item Type	Τύπος προϊόντος
11.	Item SubType	Υπο-τύπος προϊόντος

Πίνακας 3-1: Πληροφορίες πηγαίου αρχείου συνόλου δεδομένων.

Ο Πίνακας 3-2 που ακολουθεί, περιλαμβάνει δείγμα τυχαίων πέντε (5) εγγραφών του αρχείου «*E-Sales Reviews_v6.xlsx*¹».

¹ https://drive.google.com/drive/folders/1arhTwOtEtactgQFV_dMtG23aSWEal7wh?usp=sharing

66	862	60	Super cute and unique top	Just received this in the mail, tried it on and am smitten. i'm usually a l, but sometimes i'm a xl (if no stretch), in retailer tops. i bought this one in l and i'm sure glad i did. very flowy, stretchy and comfortable. i also bought the meda lace top from one september and they are very similar expect this is more of a t-shirt and the other is more of a blouse. i almost think i could've gotten a m in this because there is a lot of extra fabric at the chest which is usually never the issue for me	4	1	9	large	top	Knits
1	1080	34		Love this dress! it's sooo pretty. i happened to find it in a store, and i'm glad i did bc i never would have ordered it online bc it's petite. i bought a petite and am 5'8". i love the length on me- hits just a little below the knee. would definitely be a true midi on someone who is truly petite.	5	1	4	large	dress	Dresses
2	1077	60	Some major design flaws	I had such high hopes for this dress and really wanted it to work for me. i initially ordered the petite small (my usual size) but i found this to be outrageously small. so small in fact that i could not zip it up! i reordered it in petite medium, which was just ok. overall, the top half was comfortable and fit nicely, but the bottom half had a very tight under layer and several somewhat cheap (net) over layers. imo, a major design flaw was the net over layer sewn directly into the zipper - it c	3	0	0	large	dress	Dresses
169	1104	47	Antoher beautiful maeve dress!	I love the dresses by maeve and this one is no exception. i was pleasantly surprised upon receiving this dress because i wasn't sure about the print size and colors when i saw it online, but the print is beautiful. it's a very simple but elegant style and the ruffle on the back adds an extra touch of interest. i like that it has enough coverage at the top (front and back), the underarms do not fall very low, and it is lined. the a-line skirt will flatter most body types. the fabric and overall q	5	1	3	large	dress	Dresses
317	844	53	Lovely printed blouse	I just purchased this beautiful printed blouse in the pink color and love it! i almost always wear a size small at retailer (34d-27-35) and the fit and length are both perfect on me. if you are smaller chested you can easily go down a size. i absolutely had to have this whe i first saw it at the store and noticed how popular it was as i had to order it due to it selling out like hot cakes there. what i like about it is the texture and the ruffles at the front plus the length of the sleeves stop ri	5	1	4	small	top	Blouses

Πίνακας 3-2: Δείγμα πέντε (5) τυχαίων εγγραφών από το σύνολο δεδομένων.

3.5 Εργαλεία ανάπτυξης εφαρμογής οπτικοποίησης στην Python

Η γλώσσα προγραμματισμού Python αναπτύχθηκε από τον Guido van Rossum, ο οποίος ξεκίνησε την υλοποίησή της το 1989. Πρόκειται για σχετικά απλή και εύκολα κατανοητή γλώσσα προγραμματισμού, η οποία οφείλει το όνομά της όχι στο γνωστό φίδι (πύθωνας) αλλά σε κωμική τηλεοπτική εκπομπή (Monty Python's Flying Circus). Τα βασικά χαρακτηριστικά της γλώσσας Python παρουσιάζονται στην Εικόνα 3-2 και συνοψίζονται στα ακόλουθα:

- 1) Αξιοπιστία: Πρόκειται για πραγματικά αξιόπιστη γλώσσα με πολύ «περιεκτικές» εντολές, που πραγματικά εκτελούν με αξιόπιστο τρόπο τις αναμενόμενες λειτουργίες.
- 2) Εύκολη εκμάθηση: Η δομή και λειτουργία της Python είναι σχετικά απλή, ενώ οι εντολές της είναι υψηλού επιπέδου και ιδιαίτερα εκφραστικές. Ως εκ τούτου, η κατανόηση και εκμάθησή της μπορεί να γίνει γρήγορα και σχετικά εύκολα.
- 3) Δουλεύει σε πολλαπλές πλατφόρμες (Cross Platform): Η Python μπορεί να λειτουργήσει σε όλα τα δημοφιλή λειτουργικά συστήματα (Mac, Windows, Linux, Unix κ.λπ.) χωρίς διαφοροποιήσεις και προβλήματα. Αυτό το καθιστά γλώσσα με υψηλό βαθμό φορητότητας.
- 4) Ανοιχτού κώδικα (Open Source): Πρόκειται για γλώσσα προγραμματισμού ανοιχτού κώδικα με όλα τα πλεονεκτήματα που αυτό δίνει στους προγραμματιστές που τη χρησιμοποιούν.
- 5) Μεγάλη ποικιλία βιβλιοθηκών: Υπάρχει μια μεγάλη ποικιλία standard βιβλιοθηκών με προετοιμασμένο κώδικα και σημαντικές λειτουργίες, που μπορούν να υλοποιηθούν χωρίς να απαιτείται συγγραφή ευρείας κλίμακας κώδικα.
- 6) Χωρίς κόστος: Η γλώσσα Python παρέχεται δωρεάν για λήψη και χρήση, γεγονός που την καθιστά ακόμη πιο ελκυστική.
- 7) Υποστηρίζει διαχείριση εξαιρέσεων (Exception Handling): Το συγκεκριμένο χαρακτηριστικό αυξάνει τις ανοχές του εκτελέσιμου κώδικα καθώς εκτελείται ακόμη και όταν υπάρχουν σφάλματα, τα οποία μπορεί ο προγραμματιστής να τα χειριστεί σε επόμενο στάδιο της ανάπτυξης και αποσφαλμάτωσης του κώδικα που συγγράφει.

- 8) Αυτόματη διαχείριση μνήμης: Η Python υποστηρίζει αυτόματη διαχείριση μνήμης, το οποίο πρακτικά σημαίνει ότι η μνήμη εκκαθαρίζεται και απελευθερώνεται αυτόματα, χωρίς να χρειάζεται παρέμβαση του προγραμματιστή.
- 9) Προηγμένα χαρακτηριστικά: Υποστηρίζει γεννήτριες (generators) και λίστες κατανόησης (lists comprehensions).



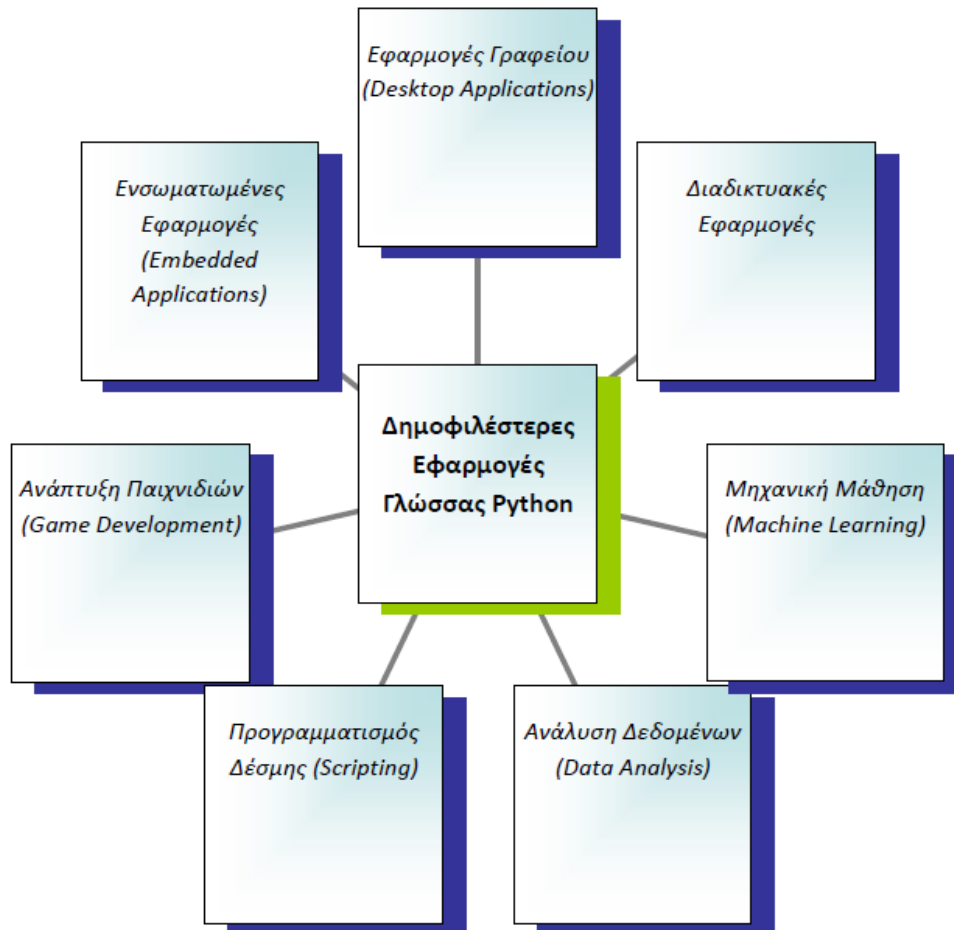
Εικόνα 3-2: Βασικά χαρακτηριστικά της γλώσσας Python.

Υπάρχουν πολλαπλές περιοχές στις οποίες ο προγραμματισμός με τη γλώσσα Python βρίσκει εφαρμογή. Η παρακάτω λίστα είναι ενδεικτική, ενώ η Εικόνα 3-3 συνοψίζει τις περιοχές αυτές:

- 1) Ανάπτυξη διαδικτυακών εφαρμογών: Δημοφιλή διαδικτυακά frameworks όπως τα Django και Flask έχουν αναπτυχθεί πάνω σε Python. Με τη βοήθειά τους μπορεί να γραφτεί

κώδικας από την πλευρά του server για λειτουργίες όπως η διαχείριση της βάσης δεδομένων, η λογική του back end, κ.λπ.

- 2) Μηχανική μάθηση (Machine Learning): Υπάρχουν πολλές εφαρμογές μηχανικής μάθησης γραμμένες σε Python. Παραδείγματα συνιστούν οι συστάσεις προϊόντων σε δημοφιλείς ιστότοπους όπως το Amazon, το eBay κ.λπ. και η αναγνώριση φωνής ή προσώπου (voice-facial recognition).



Εικόνα 3-3: Δημοφιλέστερες εφαρμογές της γλώσσας Python.

- 3) Ανάλυση δεδομένων (Data Analysis): Η ανάλυση και οπτικοποίηση δεδομένων με χρήση διαδραστικών γραφημάτων είναι ιδιαίτερα δημοφιλής χρησιμοποιώντας την Python. Το συγκεκριμένο πεδίο εφαρμογής είναι και αυτό στο οποίο εμπίπτει η παρούσα μελέτη περίπτωσης.

- 4) Προγραμματισμός δέσμης (Scripting): Πρόκειται για μικρά προγράμματα που εν γένει χρησιμοποιούνται για την αυτοματοποίηση απλών εργασιών όπως για παράδειγμα η αποστολή αυτοματοποιημένων μηνυμάτων ηλεκτρονικού ταχυδρομείου. Τέτοιου τύπου εφαρμογές χρησιμοποιούν πολλές φορές τη γλώσσα προγραμματισμού Python.
- 5) Ανάπτυξη παιχνιδιών (Game Development): Πρόκειται για περιοχή στην οποία επίσης γίνεται όλο και δημοφιλέστερη η χρήση της Python.
- 6) Ενσωματωμένες εφαρμογές (Embedded Applications): Η ανάπτυξη τέτοιου τύπου εφαρμογών με χρήση της Python, σταδιακά αυξάνεται.
- 7) Εφαρμογές γραφείου (Desktop Applications): Εφαρμογές γραφείου με χρήση της Python έχουν ήδη καταστεί δημοφιλείς, κάνοντας χρήση βιβλιοθηκών όπως είναι η Tkinter και η QT.

Όπως ήδη αναφέρθηκε, ένα από τα σημαντικά χαρακτηριστικά πλεονεκτήματα της Python είναι η ευρύτατη γκάμα βιβλιοθηκών, που περιέχουν σημαντική λειτουργικότητα έτοιμη προς χρήση, χωρίς να απαιτείται η συγγραφή εκτενούς κώδικα από τον προγραμματιστή. Ειδικότερα, στις λειτουργίες οπτικοποίησης δεδομένων, που ενδιαφέρουν ιδιαίτερα στο πλαίσιο της παρούσας εργασίας, ξεχωρίζουν ορισμένες βιβλιοθήκες οι οποίες αξίζει να αναφερθούν, λαμβάνοντας υπόψη το γεγονός ότι χρησιμοποιούνται και στη μελέτη περίπτωσης που έχει αναπτυχθεί και αφορά στην οπτικοποίηση της ανάλυσης συναισθημάτων.

Καταρχάς, η Matplotlib, παραδοσιακή βιβλιοθήκη σχεδίασης της Python, εξακολουθεί να έχει ισχυρά λειτουργικά χαρακτηριστικά και η χρήση της είναι κατά κανόνα δεδομένη για εξειδικευμένες βιβλιοθήκες όπως τα NetworkX και Pandas data frames. Από την άλλη, η Matplotlib υστερεί σε κάποια σημεία όπως είναι οι δυνατότητες ολοκλήρωσης (integration capabilities) και η διαδραστικότητα. Στα σημεία αυτά η Bokeh και η Plotly υπερτερούν και έχουν το κορυφαίο πλεονέκτημα ότι σχεδιαστικά είναι χτισμένα πάνω σε Javascript, γεγονός που καθιστά τη δυνατότητα μετατροπής γραφικών από σε Javascript εφόσον απαιτηθεί, εύκολη υπόθεση.

Ειδικότερα η Plotly, η οποία αποτελεί και την καρδιά της παρούσας υλοποίησης, έχει ένα ακόμη συγκριτικό πλεονέκτημα αφού είναι ολοκληρωμένη με το Dash, ένα framework το οποίο δίνει τη δυνατότητα απευθείας κλήσης της λειτουργικότητας από web browser χωρίς πρακτικά να

απαιτείται προγραμματισμός έξω από την Python. Αυτό συμβαίνει γιατί ο απαιτούμενος κώδικας Javascript, συμπεριλαμβανομένων των στοιχείων της React και των απαιτούμενων APIs δημιουργούνται απευθείας μέσα από την Dash, καθιστώντας έτσι τις λειτουργίες διαδραστικότητας πλήρεις και την υλοποίησή τους απλή υπόθεση. Τέλος, χρησιμοποιώντας την Dash, και συνδυάζοντάς την με το open source framework Bootstrap, καθίσταται εφικτή η εύκολη και γρήγορη διαμόρφωση του απαιτούμενου αισθητικά στυλ της εφαρμογής που υλοποιείται.

3.6 Παρουσίαση πρωτοτύπου (Prototype) υλοποίησης

Για την ανάπτυξη και εκτέλεση της παρούσας υλοποίησης, έχει χρησιμοποιηθεί το Python toolset “Anaconda3” και ειδικότερα το Jupyter Notebook, μέσω του οποίου χτίστηκε σταδιακά η εφαρμογή, αλλά και το “Powershell Prompt”, το οποίο αποτελεί το σημείο εκτέλεσης του τελικού κώδικά της εφαρμογής. Στην ενότητα αυτή επεξηγείται η δομή και το περιεχόμενο της Python εφαρμογής που έχει αναπτυχθεί.

Τα επιμέρους αρχεία² που δομούν την πολυσέλιδη εφαρμογή οπτικοποίησης συναισθημάτων, η οποία είναι προσβάσιμη μέσω web browser, παρουσιάζονται και επεξηγούνται στον Πίνακα 3-3.

² <https://drive.google.com/drive/folders/1seny9BZ8kolwzrTeXXt5CIP64CYhPrTx?usp=sharing>

	Αρχείο	Λειτουργικότητα
1.	index.py	Ένα από τα σημαντικότερα αρχεία της εφαρμογής, αφού αποτελεί το συνδετικό κρίκο των επιμέρους σελίδων της. Επίσης, υλοποιεί το βασικό μενού επιλογών της εφαρμογής, που εμφανίζεται πάνω δεξιά σε κάθε σελίδα της και διασφαλίζει την πλοήγηση από σελίδα σε σελίδα.
2.	app.py	Αρχείο που διασφαλίζει την κλήση των επιμέρους τμημάτων της εφαρμογής και επίσης καθορίζει το στυλ της (bootstrap theme).
3.	E-Sales Reviews_v6.xlsx	Αρχείο που περιλαμβάνει το σύνολο δεδομένων πάνω στο οποίο πραγματοποιείται η ανάλυση.
4.	/assets/sentiment-analysis.png	Εικόνα που χρησιμοποιείται για εικαστική βελτίωση της εφαρμογής και απεικονίζεται πάνω αριστερά σε κάθε σελίδα της.
5.	/apps/__init__.py	Το συγκεκριμένο αρχείο είναι κενό, η ύπαρξή του όμως είναι απαραίτητη για την ορθή λειτουργία της εφαρμογής.
6.	/apps/sent_visual_v13a.py	Αρχείο που περιλαμβάνει τον απαιτούμενο Python κώδικα για την υλοποίηση της πρώτης σελίδας της εφαρμογής (Data Sample Basics).
7.	/apps/sent_visual_v13b.py	Αρχείο που περιλαμβάνει τον απαιτούμενο Python κώδικα για την υλοποίηση της δεύτερης σελίδας της εφαρμογής (Visualizations).
8.	/apps/sent_visual_v13c.py	Αρχείο που περιλαμβάνει τον απαιτούμενο Python κώδικα για την υλοποίηση της τρίτης σελίδας της εφαρμογής (Box Plots).
9.	/apps/sent_visual_v13d.py	Αρχείο που περιλαμβάνει τον απαιτούμενο Python κώδικα για την υλοποίηση της τέταρτης σελίδας της εφαρμογής (Latent Analysis).

Πίνακας 3-3: Τα αρχεία της εφαρμογής και η λειτουργικότητά τους.

Για να εκκινήσει η εκτέλεση της εφαρμογής πρέπει να εκτελεστεί ο κώδικας του αρχείου «*index.py*». Μετά την εκτέλεση του εν λόγω κώδικα ανοίγει αυτόματα ο περιηγητής ιστοσελίδων (web browser) στη σελίδα «*127.0.0.1*», η οποία φιλοξενεί την εφαρμογή που έχει αναπτυχθεί και παρέχει τη δυνατότητα αλληλεπίδρασης με τον τελικό χρήστη.

3.7 Επεξήγηση – Ανάλυση Αποτελεσμάτων

Κατόπιν εκτέλεσης του κώδικα της εφαρμογής, όπως εξηγήθηκε στην προηγούμενη ενότητα, εμφανίζεται στο πάνω μέρος της σελίδας του περιηγητή η αρχική μπάρα με τον τίτλο και τις βασικές πληροφορίες της, και το μενού πλοήγησης, όπως φαίνεται στο επάνω τμήμα της Εικόνας 3-4. Αμέσως πιο κάτω εμφανίζονται βασικές πληροφορίες για την Εφαρμογή, οι οποίες μάλιστα επαναλαμβάνονται και στο κάτω μέρος των επόμενων σελίδων. Κατόπιν επιλογής από το

χρήστη της σελίδας «*Data Sample Basics*», εμφανίζονται και τα υπόλοιπα στοιχεία της συγκεκριμένης σελίδας, τα οποία είναι τα ακόλουθα:

- 1) Ένα δείγμα πέντε (5) τυχαία επιλεγμένων εγγραφών του συνόλου δεδομένων που χρησιμοποιείται. Όπως φαίνεται στην Εικόνα 3-4, ο χρήστης μπορεί να δει όλα τα πεδία ή μόνο έναν αριθμό από αυτά, επιλέγοντας αντίστοιχα στις επιλογές που παρέχονται στο πάνω μέρος του Πίνακα «*Πλήρης Πίνακας*» (Full Table) ή «*Τμήμα του Πίνακα*» (Condensed Table).
- 2) Στο κάτω μέρος των απεικονιζόμενων δεδομένων, παρέχεται σύνδεσμος (link) που μπορεί να κατευθύνει το χρήστη στο site από το οποίο έχει αντληθεί το αρχείο δεδομένων, ώστε να λάβει περισσότερες πληροφορίες και στατιστικά στοιχεία σχετικά με αυτό. Ο σύνδεσμος είναι <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews> και οδηγεί στο χρησιμοποιούμενο αρχείο δεδομένων με τίτλο «*Women's E-Commerce Clothing Reviews*», είναι μορφής .csv, το οποίο περιλαμβάνει κριτικές πελατών που έχουν πραγματοποιήσει αγορές ρούχων.
- 3) Στο κατώτερο τμήμα της αρχικής σελίδας εμφανίζονται βασικά συγκεντρωτικά στατιστικά του συνόλου δεδομένων, ώστε να παρέχεται μία συνολική αρχική εικόνα των δεδομένων. Ειδικότερα, εμφανίζονται ο Αριθμός Κριτικών (Number of Reviews), ο οποίος είναι ίσος με 20.161 όσες και οι συνολικές εγγραφές του αρχείου δεδομένων, η Μέση Ηλικία Πελατη (Average Customer Age), που είναι ίση με 43,17 έτη, και το Μέσο Πλήθος Λέξεων ανά κριτική (Average Word Count per Review), το οποίο και ισούται με 58.49 λέξεις.

Επιλέγοντας από το μενού πλοήγησης τη σελίδα «*Visualizations*» εμφανίζεται η δεύτερη σελίδα της εφαρμογής, η οποία περιλαμβάνει δύο μεγάλες ομάδες γραφικών. Η πρώτη ομάδα δίνει τη δυνατότητα απεικόνισης της κατανομής των αξιολογήσεων με βάση έξι (6) διαφορετικά κριτήρια ανάλογα με την επιλογή της αντίστοιχης καρτέλας (tab) από το χρήστη. Τα κριτήρια αυτά είναι τα ακόλουθα:

- 1) ανά Σκορ Αξιολόγησης (by Review Rating)
- 2) ανά Πολικότητα (by Polarity)
- 3) ανά Ηλικία (by Age)
- 4) ανά Μήκος Κειμένου Αξιολόγησης (by Text Length)

- 5) ανά Πλήθος Λέξεων Κειμένου Αξιολόγησης (by Text Word Count)
- 6) ανά Δείκτη Σύστασης (by Recommended Indicator)

Η Εικόνα 3-5 δείχνει την κατανομή των κριτικών με βάση το σκορ αξιολόγησης (by Review Rating). Παρατηρεί κανείς εύκολα ότι οι περισσότερες κριτικές έχουν στην πλειοψηφία τους υψηλά σκορ με κυρίαρχες τις βαθμολογίες το τέσσερα (4) και κυρίως το πέντε (5). Στην Εικόνα 3-6 απεικονίζεται η κατανομή των κριτικών με βάση την πολικότητα. Εδώ παρατηρείται μια συγκέντρωση των κριτικών σε χαμηλά και κυρίως μεσαία θετικά επίπεδα πολικότητας και αυτό γίνεται εύκολα αντιληπτό από την πυκνότητα του γραφήματος στα μεσαία τμήματά του (πολικότητες μεταξύ 0 και 0.6). Στην Εικόνα 3-7 φαίνεται η κατανομή των κριτικών με βάση την ηλικία των πελατών. Εύκολα παρατηρεί κανείς ότι κυριαρχούν οι ηλικίες μεταξύ 32 και 48 ετών, ενώ οι πολύ μικρές και οι μικρές και οι πολύ μεγάλες ηλικίες είναι σπάνιες, όπως αναμενόταν. Η Εικόνα 3-8 δείχνει την κατανομή των κριτικών με βάση το μήκος του κειμένου. Υπάρχει όπως φαίνεται ποικιλία μηκών, αλλά και συγκέντρωση σημαντικού αριθμού κριτικών στο δεξί μέρος του γραφήματος όπου απεικονίζονται κείμενα μήκους περίπου πεντακοσίων (500) χαρακτήρων. Στο ραβδόγραμμα της Εικόνας 3-9 παρουσιάζεται η κατανομή των κριτικών με βάση τον αριθμό των λέξεων που χρησιμοποιήθηκαν. Παρατηρείται μία σχετικά ομοιόμορφη κατανομή στο εύρος μεταξύ δέκα (10) και ογδόντα (80) λέξεων, και μία περισσότερο έντονη συγκέντρωση στο εύρος των εκατό (100) λέξεων. Το τελευταίο γράφημα αυτής της Ομάδας (Εικόνα 3-10) παρουσιάζεται η κατανομή των κριτικών σε σχέση με το δείκτη σύστασης. Μπορεί εύκολα κανείς να διαπιστώσει ότι περίπου των 80% των κριτικών συστήνουν το υπό αξιολόγηση προϊόν (δείκτης σύστασης = 1), ενώ το 20% των κριτικών τοποθετούνται αρνητικά (δείκτης σύστασης = 0).

Η δεύτερη ομάδα γραφημάτων της σελίδας «Visualizations» ξεκινάει με μία λίστα πολλαπλών επιλογών/ φίλτρων, που δίνουν τη δυνατότητα στο χρήστη να επιλέξει κριτήρια-φίλτρα, βάσει των οποίων θα απεικονιστούν τα τρία γραφήματα που ακολουθούν. Τα συγκεκριμένα φίλτρα εφαρμόζονται συνδυαστικά, αντανακλώνονται δυναμικά στα γραφήματα και είναι τα ακόλουθα:

- 1) Το μέγεθος του ρούχου που έχει αξιολογηθεί. Διατίθενται τρεις (3) επιλογές, οι οποίες είναι το κανονικό (normal), το μεγάλο (large) και το μικρό (small) μέγεθος.

- 2) Το ηλικιακό εύρος του πελάτη που έχει κάνει την αξιολόγηση (age). Διατίθενται τρεις (3) επιλογές, οι οποίες είναι το ηλικιακό εύρος κάτω των 35 (age < 35), το ηλικιακό εύρος μεταξύ 35 και 54 (age 35-54) και το ηλικιακό εύρος άνω των 54 (age > 54).
- 3) Το σκορ αξιολόγησης. Διατίθενται πέντε (5) επιλογές, οι οποίες είναι τα σκορ ένα (1), δύο (2), τρία (3), τέσσερα (4) και πέντε (5).
- 4) Ο τύπος ρούχου που έχει αξιολογηθεί. Διατίθενται έξι (6) επιλογές, οι οποίες είναι οι τύποι αξεσουάρ (accessories), bottom, φόρεμα (dress), jacket, top και μοντέρνα (trendy).

Όπως ήδη αναφέρθηκε, με βάση τα φίλτρα της εκάστοτε επιλογής, ενημερώνονται αυτόματα τρία (3) γραφήματα, τα οποία είναι τα εξής:

- 1) Οι δέκα (10) κορυφαίες λέξεις εξαιρουμένων των κοινών λέξεων (stop words) επί του επιλεγθέντος δείγματος βάσει των φίλτρων.
- 2) Τα δέκα (10) κορυφαία τρίλεξα εξαιρουμένων των κοινών λέξεων (stop words) επί του επιλεγθέντος δείγματος βάσει των φίλτρων.
- 3) Οι δέκα (10) κορυφαίες σημάνσεις (tags) του κειμένου επί του επιλεγθέντος δείγματος βάσει των φίλτρων.

Η Εικόνα 3-11 περιλαμβάνει τις δέκα (10) κορυφαίες σε συχνότητα εμφάνισης λέξεις εξαιρουμένων των κοινών λέξεων (stop words), για μέγεθος του ρούχου κανονικό (normal), ηλικιακό εύρος του πελάτη μεταξύ 35 και 54 (age 35-54), σκορ αξιολόγησης πέντε (5) και τύπος ρούχου φόρεμα (dress). Οι κορυφαίες σε συχνότητα εμφάνισης λέξεις είναι “*dress*”, “*love*” και “*size*”, με το “*dress*” να ξεχωρίζει με εφτακόσιες είκοσι τέσσερις (724) εμφανίσεις.

Στην Εικόνα 3-12 απεικονίζονται τα δέκα (10) κορυφαία τρίλεξα εξαιρουμένων των κοινών λέξεων (stop words), για μέγεθος του ρούχου κανονικό (normal), ηλικιακό εύρος του πελάτη μεταξύ 35 και 54 (age 35-54), σκορ αξιολόγησης πέντε (5) και τύπος ρούχου φόρεμα (dress). Μπορεί κανείς να παρατηρήσει ότι τα δημοφιλέστερα τρίλεξα είναι αναφορές γύρω από το μέγεθος του ρούχου (size). Ειδικότερα, τα τρία δημοφιλέστερα τρίλεξα είναι τα “*absolutely love dress*” με έντεκα (11) εμφανίσεις, “*fits true size*” με οκτώ (8) εμφανίσεις, “*fit true size*”, “*fits like glove*” και “*love love love*” με πέντε (5) εμφανίσεις.

Στο γράφημα της Εικόνας 3-13 απεικονίζονται οι δέκα (10) κορυφαίες σημάνσεις (tags) εξαιρουμένων των κοινών λέξεων (stop words), για μέγεθος του ρούχου κανονικό (normal),

ηλικιακό εύρος του πελάτη μεταξύ 35 και 54 (age 35-54), σκορ αξιολόγησης πέντε (5) και τύπος ρούχου φόρεμα (dress). Μπορεί κανείς να παρατηρήσει ότι οι δημοφιλέστερες σημάσεις είναι τα “*NN*” με τριάντα μία (31) εμφανίσεις, “*DT*” με δέκα οκτώ (18) εμφανίσεις και “*CD*” με δέκα πέντε (15) εμφανίσεις.

Η τρίτη σελίδα της εφαρμογής περιλαμβάνει δύο (2) παραμετρικά, δυναμικά γραφήματα τύπου boxplot. Ειδικότερα, στην Εικόνα 3-14 απεικονίζεται η κατανομή πολικότητας ανά τύπο ρούχου για τους διαθέσιμους τύπους ρούχων (“*top*”, “*dress*”, “*bottom*”, “*accessories*”, “*jacket*”, “*trendy*”). Πρόκειται για διαδραστικό γράφημα, που εκτός από τη γραφική απεικόνιση, παρέχει πλήθος επιπλέον στατιστικών στοιχείων όταν το βέλος του ποντικιού ακουμπήσει πάνω στα γραφικά. Τέτοιες πληροφορίες είναι τα μέγιστο (max), ελάχιστο (min), μέσος (median), κάτω φράγμα (lower fence), άνω φράγμα (upper fence) και συντελεστές q_1 και q_3 . Παρατηρώντας και αλληλεπιδρώντας με το γράφημα της Εικόνας 3-14, μπορεί κανείς να διαπιστώσει ότι οι κατηγορίες ρούχων με τον υψηλότερο μέσο (median) πολικότητας είναι τα “*accessories*” με 0.246 και “*top*” με 0.245. Όλες οι κατηγορίες έχουν μέγιστη τιμή ίση με ένα (1), πλην της “*trendy*”, της οποίας η μέγιστη τιμή είναι ίση με 0.667, ενώ οι χαμηλότερες ελάχιστες τιμές παρατηρούνται στις κατηγορίες “*top*” (ίση με -0.975) και “*dress*” (ίση με -0.917). Ένα άλλο σημείο που μπορεί να σχολιαστεί είναι το εύρος τιμών ανά κατηγορία, το οποίο είναι μεγαλύτερο για τις κατηγορίες “*top*” και “*dress*”.

Στην Εικόνα 3-15 απεικονίζεται η κατανομή μήκους κριτικής ανά τύπο ρούχου για τους διαθέσιμους τύπους ρούχων (“*top*”, “*dress*”, “*bottom*”, “*accessories*”, “*jacket*”, “*trendy*”). Παρατηρώντας και αλληλεπιδρώντας με το γράφημα της Εικόνας 3-15, μπορεί κανείς να διαπιστώσει ότι οι κατηγορίες ρούχων με τον υψηλότερο μέσο (median) μήκους κριτικής είναι τα “*trendy*” με 346.5 και “*dress*” με 323. Οι κατηγορίες “*dress*” και “*top*” έχουν τις υψηλότερες μέγιστες τιμές (ίσες με 508), ενώ η χαμηλότερη ελάχιστη τιμή παρατηρείται στην κατηγορία “*top*” και είναι ίση με 9. Ένα άλλο σημείο που μπορεί να σχολιαστεί είναι το εύρος τιμών ανά κατηγορία, το οποίο είναι μεγαλύτερο για την κατηγορία “*top*” και ακολουθούν οι κατηγορίες “*dress*” και “*bottom*”.

Στην Εικόνα 3-16 απεικονίζεται η κατανομή πολικότητας ανά σκορ αξιολόγησης για τα διαθέσιμα σκορ αξιολόγησης (1, 2, 3, 4, 5). Παρατηρώντας και αλληλεπιδρώντας με το γράφημα της Εικόνας 3-16, μπορεί κανείς να διαπιστώσει τα σκορ αξιολόγησης με τον υψηλότερο μέσο

(median) πολικότητας είναι το “5” με 0.29 και το “4” με 0.22. Τα σκορ “1” και “2” έχουν τις μικρότερες μέγιστες τιμές, ίσες αντίστοιχα με 0.8 και 0.91, ενώ οι χαμηλότερες ελάχιστες τιμές παρατηρούνται επίσης στα σκορ “1” (ίση με -0.975) και “2” (ίση με -0.917), ενώ το σκορ “4” παρουσιάζει την υψηλότερη ελάχιστη τιμή (ίση με -0.522). Ένα άλλο σημείο που μπορεί να σχολιαστεί είναι το εύρος τιμών ανά κατηγορία, το οποίο είναι μεγαλύτερο και πάλι για τα σκορ “1” και “2”.

Τέλος, στην Εικόνα 3-17 απεικονίζεται η κατανομή της ηλικίας πελάτη ανά σκορ αξιολόγησης για τα διαθέσιμα σκορ αξιολόγησης (1, 2, 3, 4, 5). Παρατηρώντας και αλληλεπιδρώντας με το γράφημα της Εικόνας 3-17, μπορεί κανείς να διαπιστώσει ότι το σκορ αξιολόγησης με τον υψηλότερο μέσο (median) ηλικίας πελάτη είναι το “1” με 42 έτη. Τα σκορ “2” και “3” έχουν τις μικρότερες μέγιστες τιμές, ίσες με 91 έτη ενώ οι χαμηλότερες ελάχιστες τιμές παρατηρούνται στα σκορ “4” και “5” και είναι ίσες με 18 έτη. Ένα άλλο σημείο που μπορεί να σχολιαστεί είναι το εύρος τιμών ανά κατηγορία, το οποίο είναι μεγαλύτερο και πάλι για τα σκορ “4” και “5”.

Τα εντυπωσιακότερα ενδεχομένως γραφήματα απεικονίζονται στην τέταρτη και τελευταία σελίδα της εφαρμογής. Πρόκειται για δύο γραφήματα λανθάνουσας ανάλυσης, στα οποία απεικονίζονται τα έξι (6) δημοφιλέστερα τρίλεξα (triwords) των κριτικών σε δύο διαφορετικούς τύπους γραφημάτων. Η λανθάνουσα ανάλυση αποτελεί σημαντική μέθοδο επεξεργασίας φυσικής γλώσσας, με την οποία μπορούν να ανακαλυφθούν σύνολα συσχετιζόμενων όρων. Η προστιθέμενη αξία της συγκεκριμένης τεχνικής είναι ότι δίνει τη δυνατότητα να κατανοηθούν έννοιες και όχι απλά όροι. Τα βασικά χαρακτηριστικά της εν λόγω τεχνικής μπορούν να συνοψιστούν στα ακόλουθα:

- Χρησιμοποιεί αλγορίθμους, οι οποίοι λαμβάνουν υπόψη τις αλληλεξαρτήσεις μεταξύ των όρων.
- Έχει τη δυνατότητα να αντιληφθεί τη σχετικότητα μεταξύ όρων και να μετασχηματίσει την αναπαράσταση των εγγράφων, μεταβαίνοντας από το χώρο των όρων στο χώρο των εννοιών (concepts).
- Μειώνει τη διαστασιμότητα της ανάλυσης, με θετικό αντίκτυπο στο χρόνο επεξεργασίας ερωτημάτων και στην κατανόηση και ερμηνεία των αποτελεσμάτων.

Στο πρώτο γράφημα, που φαίνεται στην Εικόνα 3-18, μπορεί κανείς εύκολα να διαπιστώσει ότι το μακράν δημοφιλέστερο τρίλεξο που εμπεριέχεται στις κριτικές είναι το “*love dress size*” με 18.051 εμφανίσεις. Το δεύτερο δημοφιλέστερο τρίλεξο είναι το “*dress beautiful love*” με 772 εμφανίσεις, το τρίτο είναι το “*shirt cute love*” με 592 εμφανίσεις και το έκτο είναι το “*size true fits*” με 120 εμφανίσεις.

Στο δεύτερο εξίσου εντυπωσιακό γράφημα (Εικόνα 3-19) εμφανίζονται τα δημοφιλή θέματα σε ομάδες (clusters). Οι κουκίδες που εμφανίζονται στο γράφημα είναι χρωματισμένες με άλλο χρώμα για κάθε ομάδα, ενώ το πλήθος τους κάνει εμφανές ότι για παράδειγμα η μακράν κυρίαρχη ομάδα είναι οι κουκίδες κόκκινου χρώματος (“*love dress size*”).

Θα πρέπει να τονιστεί ότι όλα τα γραφήματα είναι διαδραστικά. Πέραν του μενού πλοήγησης στην εφαρμογή, των καρτελών (tabs) και των λιστών επιλογής, σε κάθε γράφημα υπάρχει απεικόνιση δεδομένων και σημαντικών πληροφοριών με την κίνηση του ποντικιού πάνω στα γραφικά, ενώ παρέχονται και όλες οι διαδράσεις που δίνει η βιβλιοθήκη Plotly σε συνδυασμό με την Dash, όπως έχουν περιγραφεί και επεξηγηθεί στην Ενότητα 3.3 και στην Εικόνα 3-1.

VISUALIZATION APPLICATION ON SENTIMENT ANALYSIS

USING PYTHON SCRIPTING & LIBRARIES

DATA SET SAMPLE - WOMENS CLOTHING

Condensed table
Full table

REVIEW_ID	ITEM_ID	CUSTOMER_AGE	REVIEW_TITLE	REVIEW_TEXT
0	767	33		Absolutely wonderful - silky and sexy and comfortable
13	767	44	Runs big	Bought the black xs to go under the larkspur midi dress because they didn't bother lining the skirt portion (grrrrrrrrrr).n
25	697	31	Falls flat	Loved the material, but i didnt really look at how long the dress was before i purchased both a large and a medium. im 5
28	684	53	Great shirt!!!!	I have several of goodhyouman shirts and i get so many compliments on them. especially the one that says forehead kis
34	697	39	Love this dress!	This is such a neat dress. the color is great and the fabric is super soft. i am tall so the long length was an added bonus.

<
>

See more detailed data on <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>

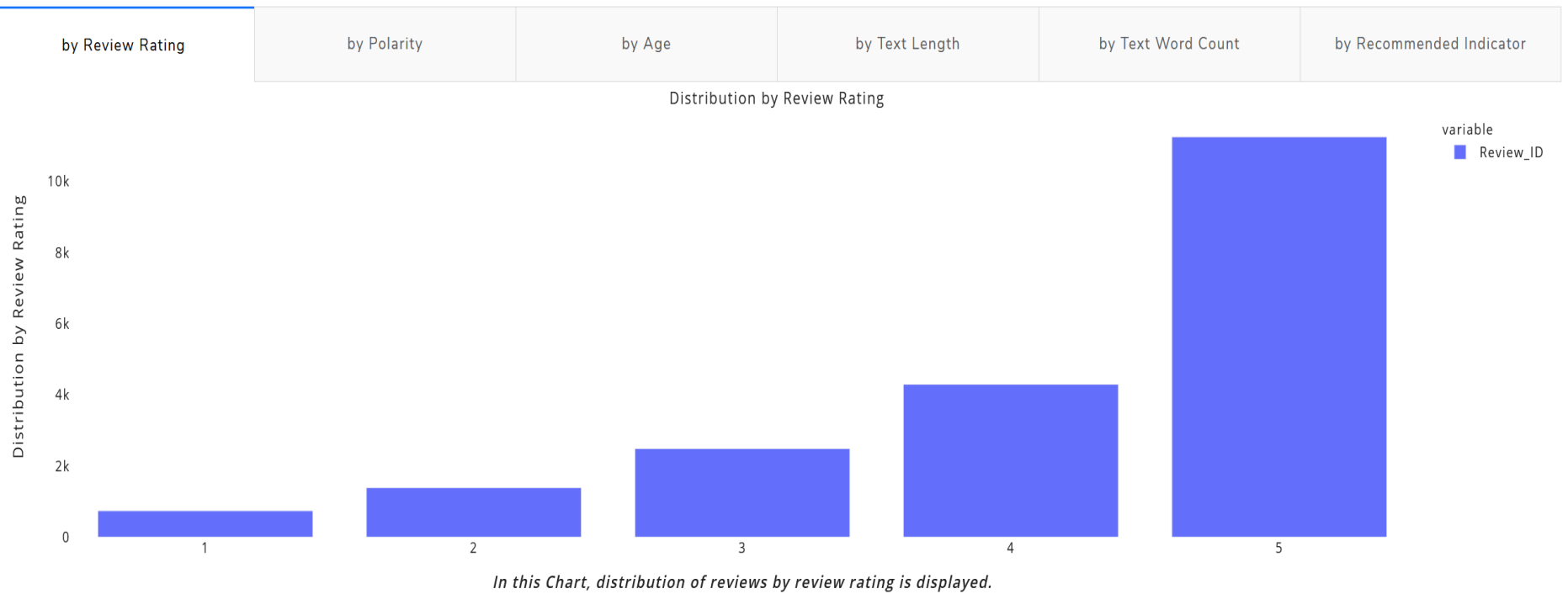
MAIN DATA SET STATISTICS	VALUE
Number of Reviews:	20161
Mean of Customer Age:	43.170725658449484
Minimum Customer Age:	18
Maximum Customer Age:	99
Standard Deviation of Customer Age:	12.270249305500188
Average Review Length:	58.48653340608105
Minimum Review Length:	2
Maximum Review Length:	114
Ratings:	[1 2 3 4 5]
Sizes:	['large' 'normal' 'small']
Item Types:	['accessories' 'bottom' 'dress' 'jacket' 'top' 'trendy']

<
>

Εικόνα 3-4: Η πρώτη σελίδα της εφαρμογής οπτικοποίησης ανάλυσης συναισθημάτων.

REPRESENTATIVE VISUALS

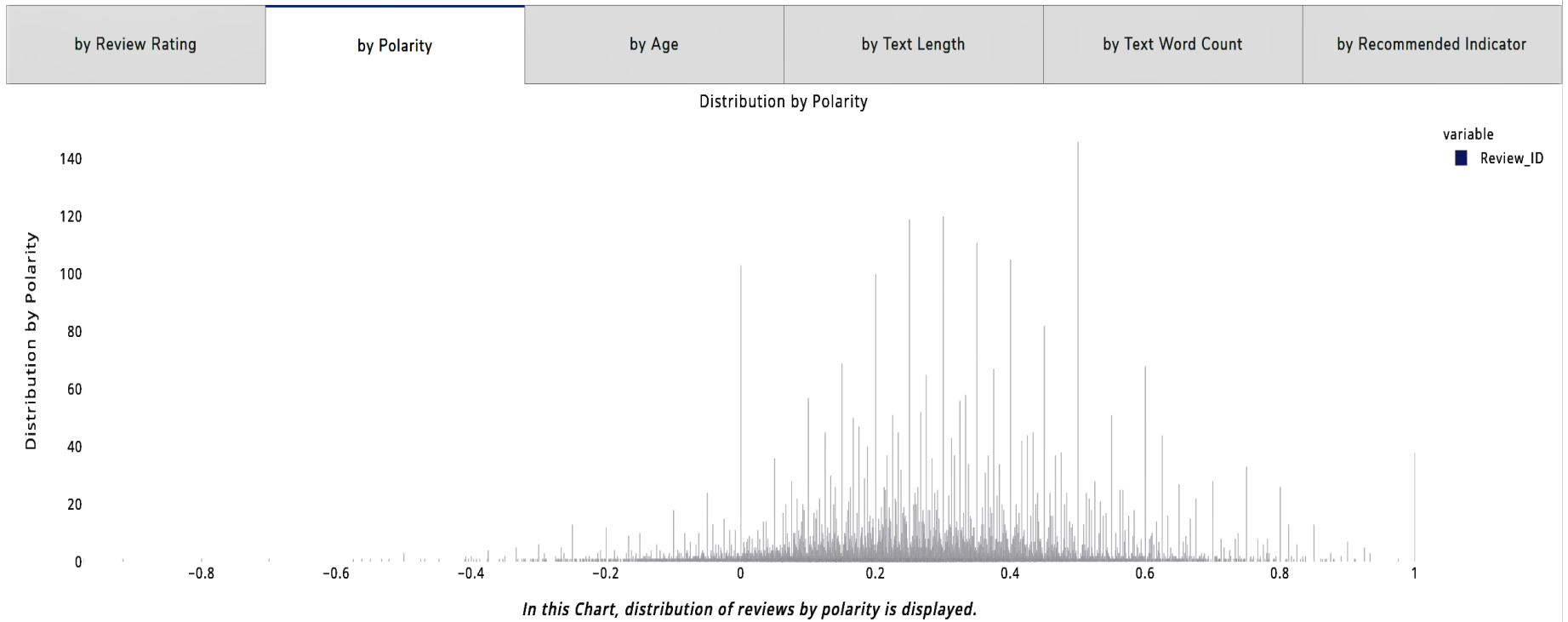
DISTRIBUTION OF REVIEWS BY SELECTED VARIABLE



Εικόνα 3-5: Κατανομή κριτικών με βάση το σκορ αξιολόγησης.

REPRESENTATIVE VISUALS

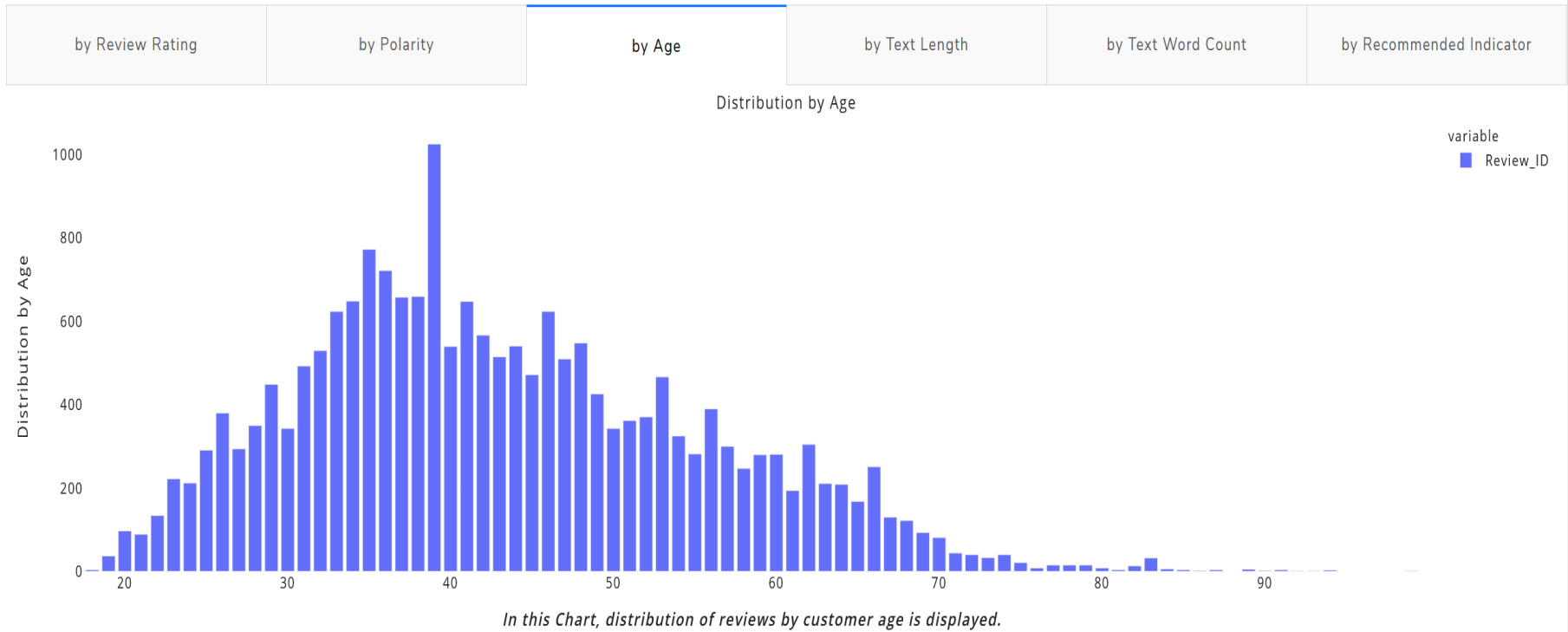
DISTRIBUTION OF REVIEWS BY SELECTED VARIABLE



Εικόνα 3-6: Κατανομή κριτικών με βάση την πολικότητα της αξιολόγησης.

REPRESENTATIVE VISUALS

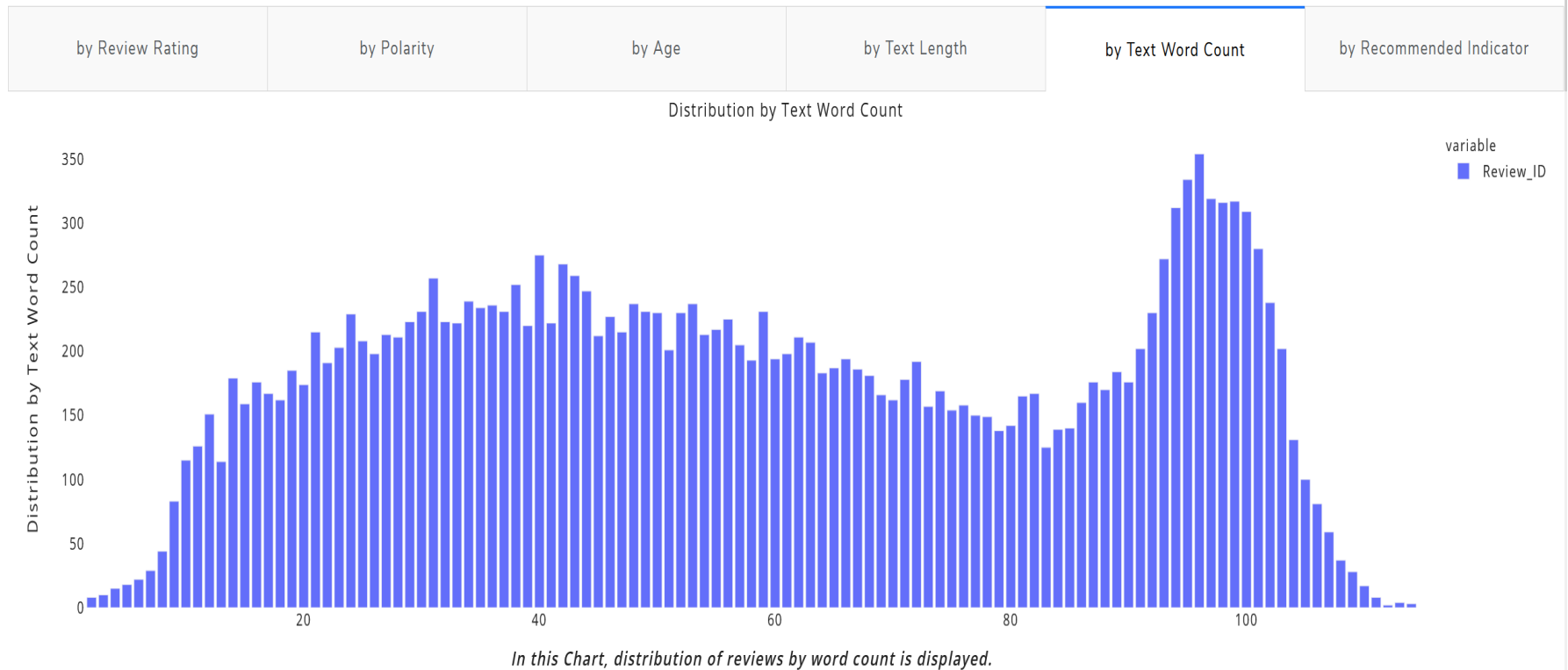
DISTRIBUTION OF REVIEWS BY SELECTED VARIABLE



Εικόνα 3-7: Κατανομή κριτικών με βάση την ηλικία του πελάτη.

REPRESENTATIVE VISUALS

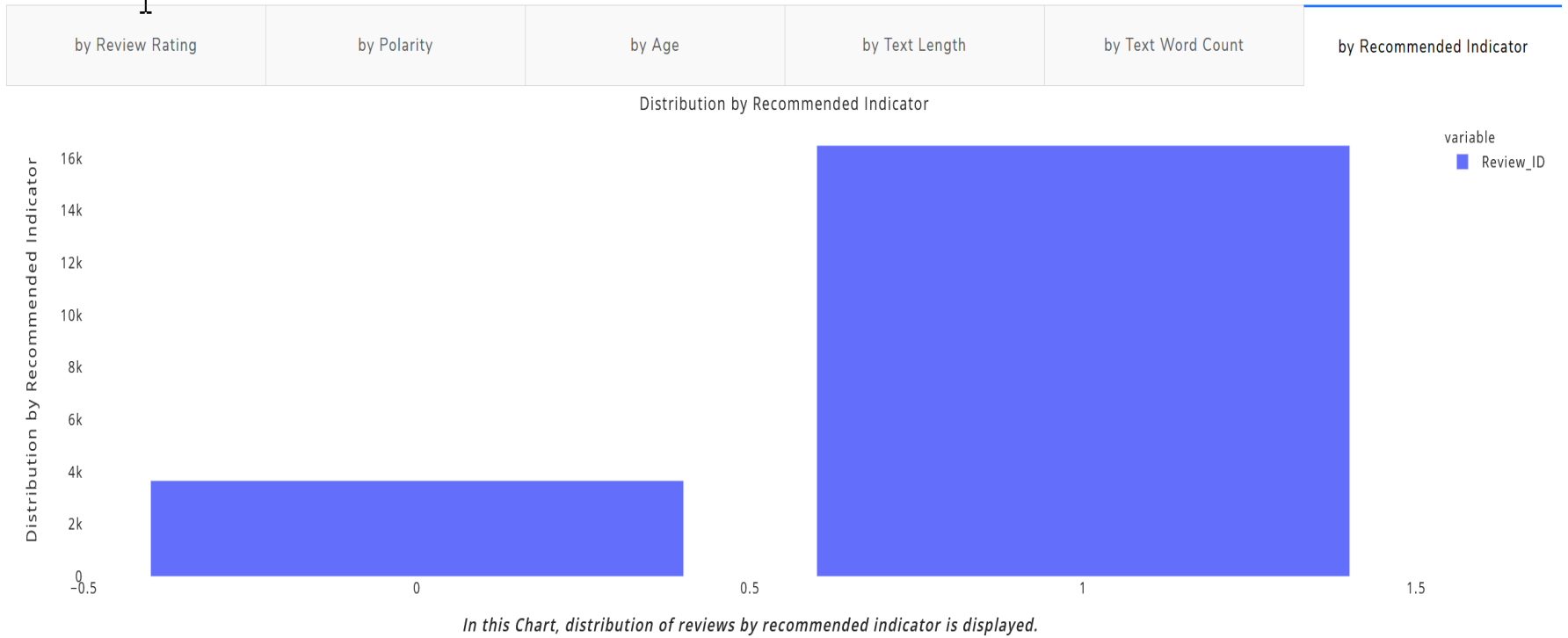
DISTRIBUTION OF REVIEWS BY SELECTED VARIABLE



Εικόνα 3-9: Κατανομή κριτικών με βάση τον αριθμό λέξεων της αξιολόγησης.

REPRESENTATIVE VISUALS

DISTRIBUTION OF REVIEWS BY SELECTED VARIABLE



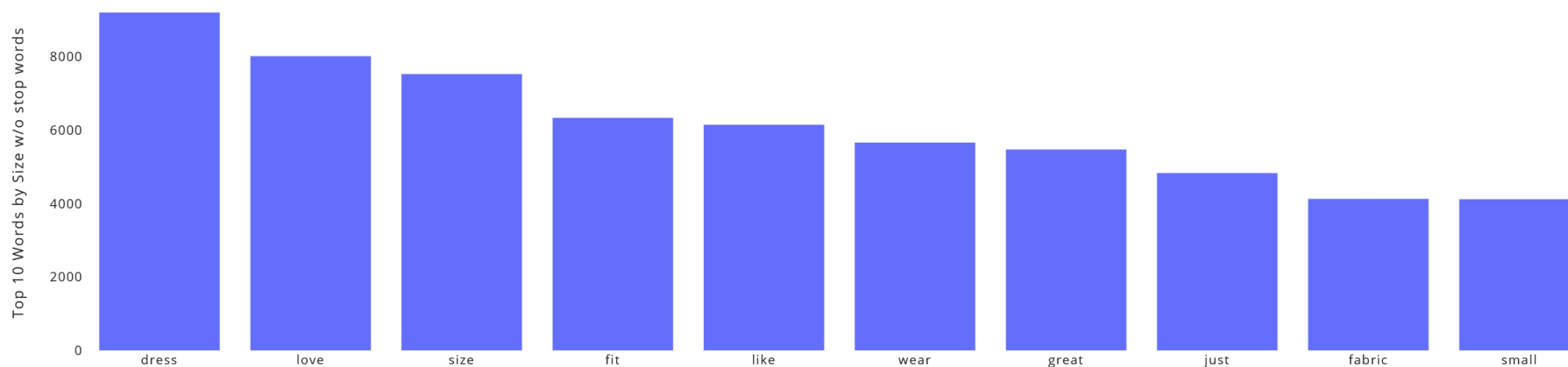
Εικόνα 3-10: Κατανομή κριτικών με βάση το δείκτη σύστασης.

Size: large
 Size: normal
 Size: small
 Recommended Indicator: 0
 Recommended Indicator: 1

Purchase Rating: 1
 Purchase Rating: 2
 Purchase Rating: 3
 Purchase Rating: 4
 Purchase Rating: 5

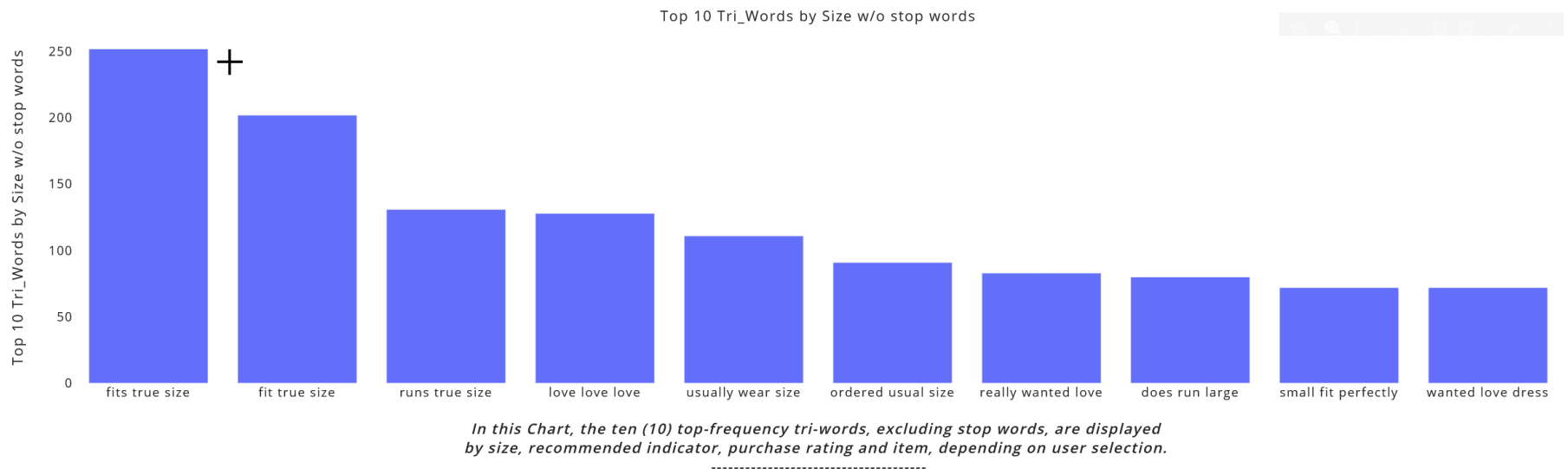
Item Type: accessories
 Item Type: bottom
 Item Type: dress
 Item Type: jacket
 Item Type: top
 Item Type: trendy

Top 10 Words by Size w/o stop words

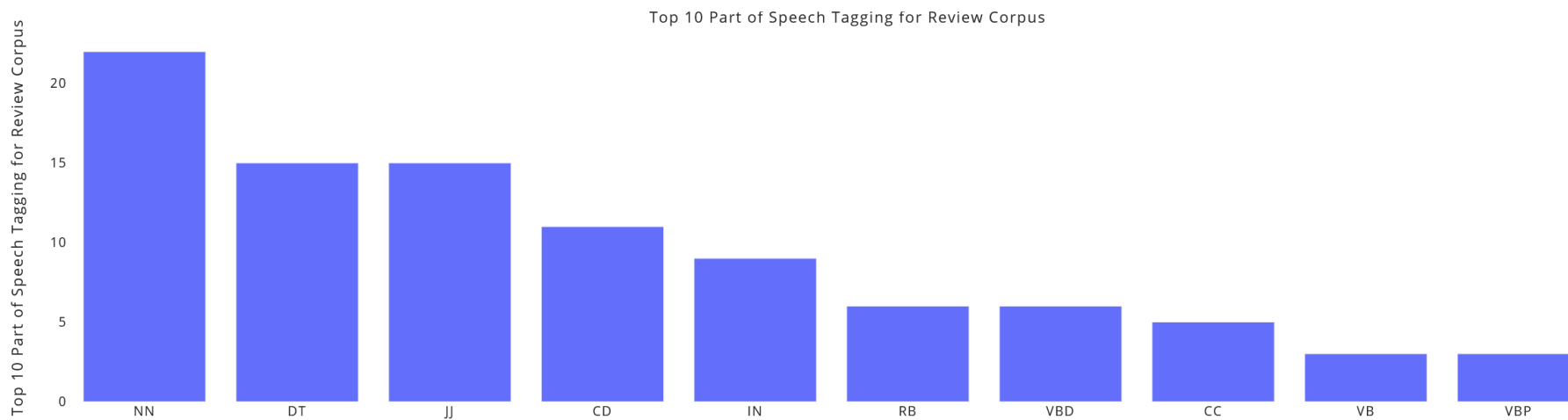


In this Chart, the ten (10) top-frequency words, excluding stop words, are displayed by size, recommended indicator, purchase rating and item depending on user selection.

Εικόνα 3-11: Οι δέκα (10) κορυφαίες λέξεις ανά μέγεθος ρούχου, δείκτη αξιολόγησης, σκορ αξιολόγησης και τύπο ρούχου, εξαιρουμένων των κοινών λέξεων (stop words).

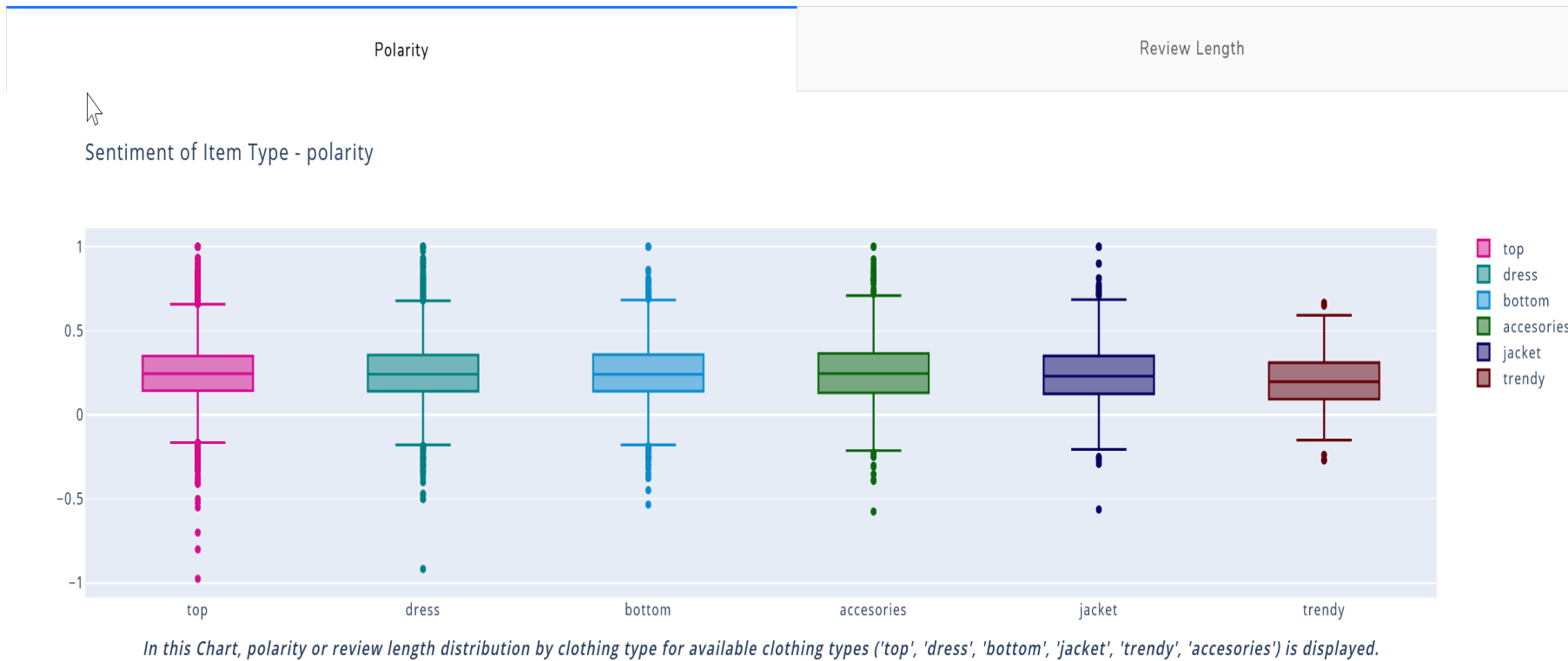


Εικόνα 3-12: Τα δέκα (10) κορυφαία τρίλεξα ανά μέγεθος ρούχου, ηλικία, σκορ αξιολόγησης και τύπο ρούχου, εξαιρουμένων των κοινών λέξεων (stop words).

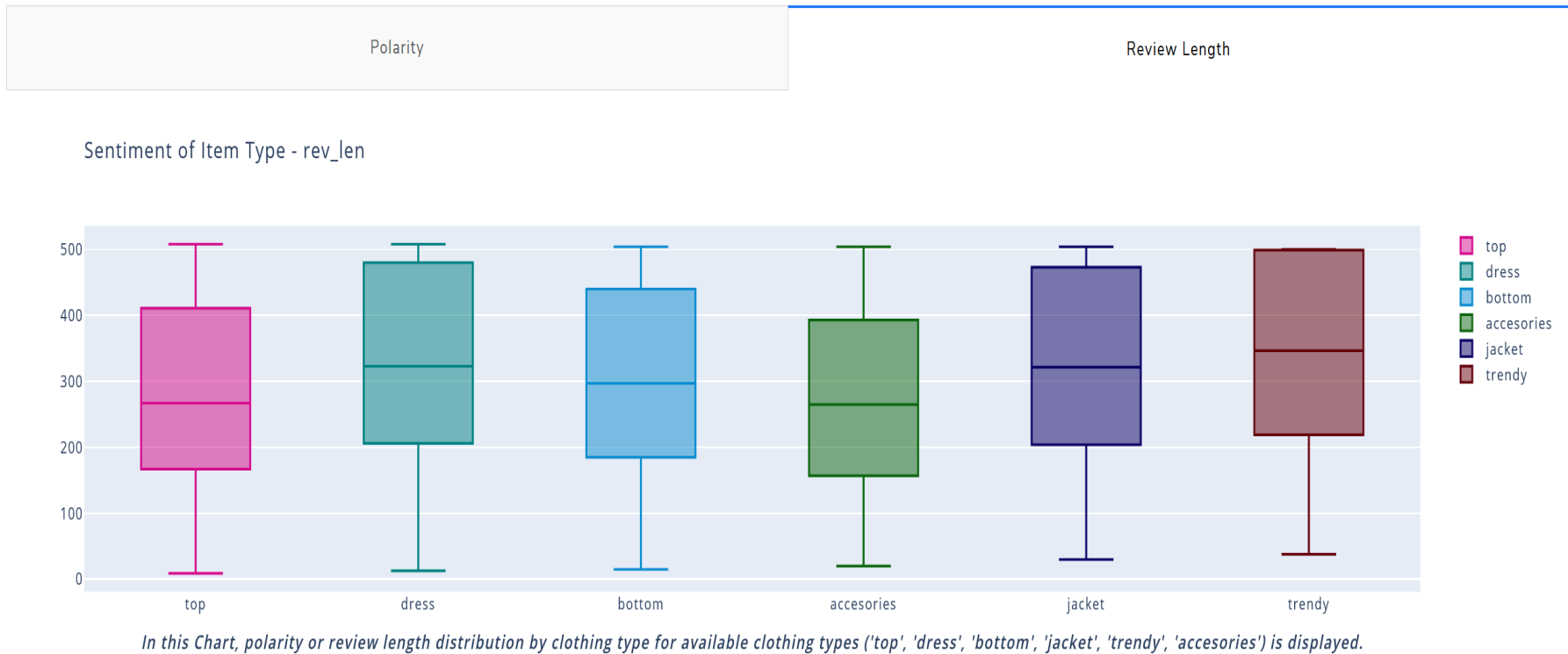


In this Chart, the ten (10) top-frequency parts of speech tagging for review corpus are displayed by size, recommended indicator, purchase rating and item, depending on user selection.

Εικόνα 3-13: Οι δέκα (10) κορυφαίες σημάνσεις (tags) του κειμένου ανά μέγεθος ρούχου, ηλικία, σκορ αξιολόγησης και τύπο ρούχου, εξαιρουμένων των κοινών λέξεων (stop words).



Εικόνα 3-14: Κατανομή πολικότητας ανά τύπο ρούχου με αναπαράσταση boxplot.



Εικόνα 3-15: Κατανομή μήκους κριτικής ανά τύπο ρούχου με αναπαράσταση boxplot.

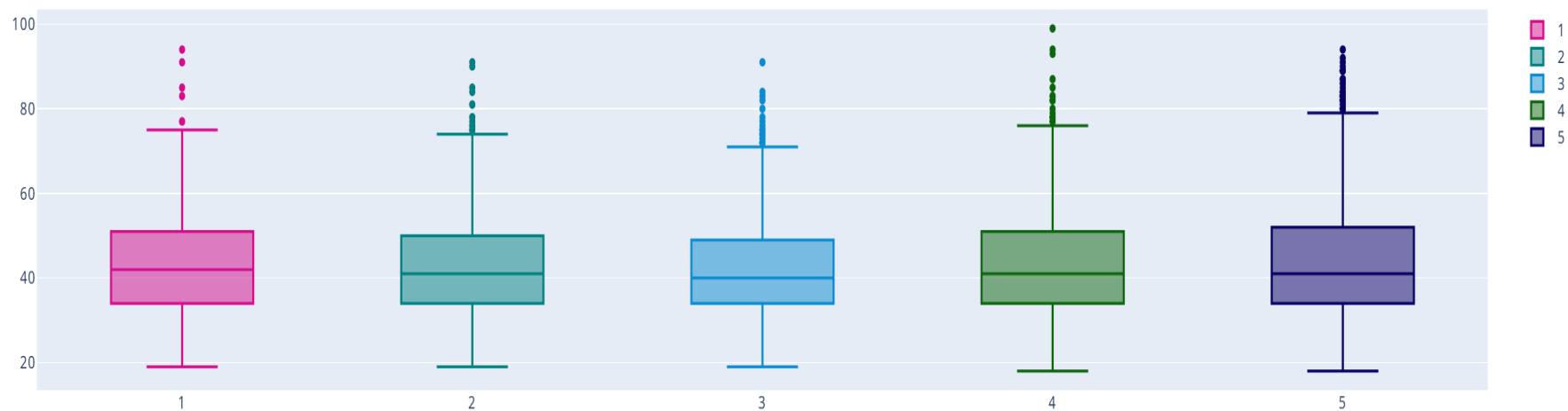


Εικόνα 3-16: Κατανομή πολικότητας ανά σκορ αξιολόγησης με αναπαράσταση boxplot.

Polarity

Customer Age

Sentiment by Purchase Rating - Customer_Age

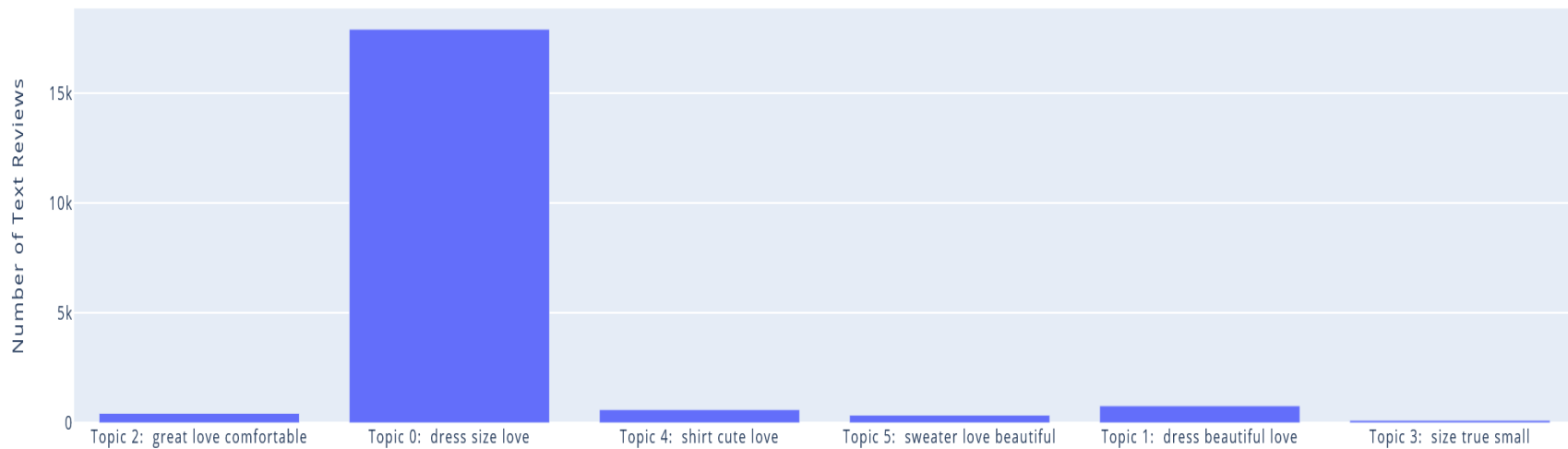


In this Chart, polarity or customer age distribution by purchase rating for all possible ratings ('1', '2', '3', '4', '5') is displayed.

Εικόνα 3-17: Κατανομή ηλικίας πελάτη ανά σκορ αξιολόγησης με αναπαράσταση boxplot.

Latent Analysis

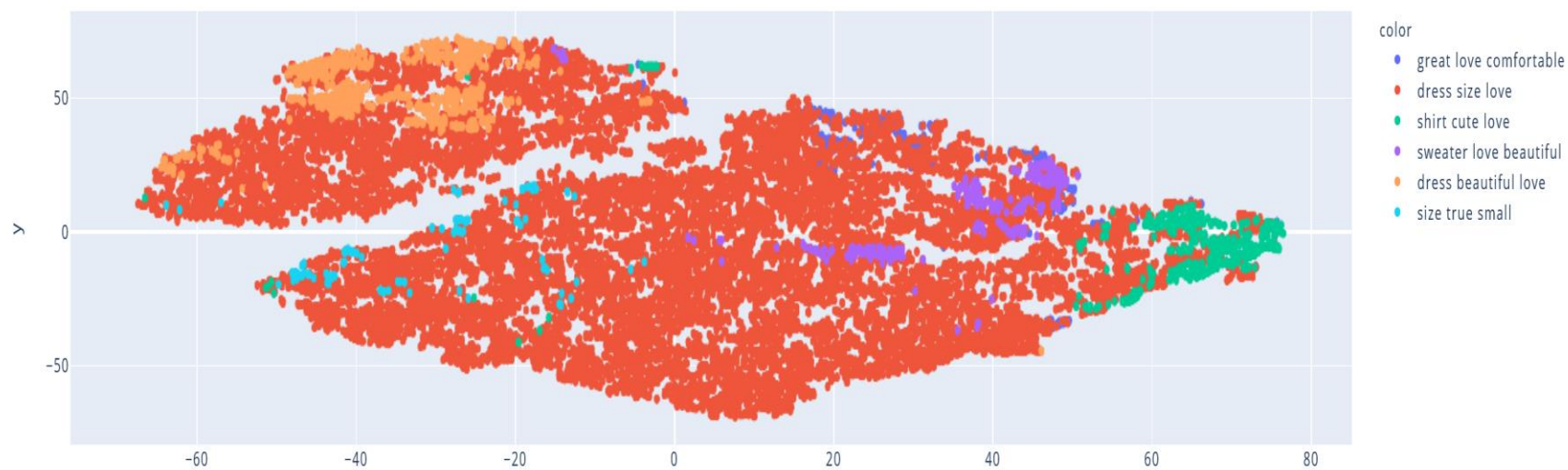
Latent Semantic Analysis Topic Counts



In this Chart, the six (6) most popular 3-word topics of the reviews are displayed based on latent analysis (NLP technique that discovers sets of related terms).

Εικόνα 3-18: Λανθάνουσα σημασιολογική ανάλυση – Έξι (6) κορυφαία θέματα.

t-SNE Clustering of 6 Latent Semantic Analysis Topics



Scatter Chart with the six (6) most popular 3-word topics of the reviews are displayed based on latent analysis.

Εικόνα 3-19: Λανθάνουσα σημασιολογική ανάλυση (TSNE clustering) – Έξι (6) κορυφαία θέματα.

4 Επίλογος

4.1 Συζήτηση αποτελεσμάτων

Στο σημείο αυτό επιχειρείται η σύνδεση των οπτικοποιήσεων, που έχουν υλοποιηθεί στο πλαίσιο της παρούσας μελέτης περίπτωσης, με τη βιβλιογραφική ανασκόπηση τεχνικών οπτικοποίησης ανάλυσης συναισθημάτων, που έχει πραγματοποιηθεί. Ειδικότερα, έπεται μία αντιστοίχιση των τεχνικών, που έχουν χρησιμοποιηθεί στο πλαίσιο της υλοποιηθείσας εφαρμογής, με αυτές που συναντώνται στη βιβλιογραφία και οι οποίες έχουν παρουσιαστεί και επεξηγηθεί στο Κεφάλαιο 2.

Δεδομένου ότι έχει υλοποιηθεί ανάλυση συναισθημάτων για την αξιολόγηση προϊόντων και σχολίων των πελατών μιας επιχείρησης, αντίστοιχη ερευνητικών πρωτοτύπων/ εργαλείων όπως είναι τα «*Affect Inspector*» (Subasic & Huettner, 2001), *Pulse* (Gamon et al., 2005), *Opinion Observer* (Liu et al., 2005), *AMAZING* (Miao et al., 2009), το παρόν έργο εμπίπτει στην κατηγορία τεχνικών ανάλυσης και οπτικοποίησης συναισθημάτων επί κριτικών/ αναφορών (Ενότητα 2.5.2).

Ως προς την πηγή προέλευσης και τις ιδιότητες των δεδομένων, η παρούσα μελέτη μπορεί υπό την ευρεία έννοια να ενταχθεί στην κατηγορία τεχνικών οπτικοποίησης ανάλυσης συναισθημάτων πάνω σε ροές δεδομένων (streams), όπως συμβαίνει και στις ερευνητικές εργασίες των Marcus και συν. (2011), Zhao και συν. (2012), Liu και συν. (2013), Kranjc και συν. (2015), Steed και συν. (2015) και Tsirakis και συν. (2016) και παρουσιάζεται στην Ενότητα 2.5.3.

Λαμβάνοντας υπόψη τις εκτελούμενες εργασίες, όπως αυτές παρουσιάζονται στην Ενότητα 2.5.4, η υλοποίηση που έχει γίνει εμπεριέχει τεχνικές, οι οποίες εμπίπτουν στην κατηγορία τεχνικών ανάλυσης πολικότητας (polarity analysis), όπως αυτές που έχουν προταθεί στις υλοποιηθέντα πρωτότυπα/ εργαλεία *SATISFI* (Taskaya & Ahmad, 2003), *Vox Civitas* (Diakopoulos et al., 2010), καθώς επίσης και αυτές των Annett και Kondrak (2008), Puri και συν. (2014), Brooks και συν. (2014) και Gamon και συν. (2008). Ακόμη, ευρύτητα στο πλαίσιο της παρούσας υλοποίησης έχει χρησιμοποιηθεί η μονοδιάστατη (aspect based) ανάλυση συναισθήματος, η οποία στη βιβλιογραφία συχνά έχει εφαρμοσθεί πάνω σε δεδομένα κριτικών πελατών. Αντίστοιχες δουλειές στη βιβλιογραφία έχουν παρουσιαστεί από τους Yatani και συν. (2011), Alper και συν. (2011), Di Caro

και Grella (2013), Gorg και συν. (2013), Wensel και Sood (2008), Cervantes και συν. (2015) και Mahmud και συν. (2016).

Επιπρόσθετα, η παρούσα υλοποίηση εμπεριέχει τεχνικές που στηρίζονται σε εργασίες αναπαράστασης και αλληλεπίδρασης, οι οποίες υποστηρίζονται απευθείας από τεχνικές απεικόνισης, όπως αυτές που παρουσιάζονται στην Ενότητα 2.5.4. Ειδικότερα, έχουν αξιοποιηθεί τεχνικές ομαδοποίησης/ κατηγοριοποίησης (clustering/ classification), οι οποίες κάνουν χρήση ετικετών (tagging), όπως αυτές που προτάθηκαν από τους Oelke και συν. (2009), Brew και συν. (2011), Kim και Lee (2014) και Sung και συν. (2016). Ακόμη, η υλοποίηση που έχει πραγματοποιηθεί, περιλαμβάνει τεχνικές συγκεντρωτικής απεικόνισης (overview), οι οποίες επιτυγχάνουν να απεικονίσουν τη «μεγάλη εικόνα», οπτικοποιώντας μεγάλο αριθμό δεδομένων ή μέσω συγκεντρωτικών (aggregated) αναπαραστάσεων, όπως αυτές που προτάθηκαν στις δουλειές των Wu και συν. (2013), Liu και συν. (2016), Miao και συν. (2009) και Scharl και συν. (2016). Επιπλέον, χρησιμοποιούνται τεχνικές πλοήγησης/ εξερεύνησης (navigation/ exploration), αφού έχει υλοποιηθεί εφαρμογή διαδραστικής πλοήγησης και εξερεύνησης των δεδομένων, όπως έχει γίνει και στις μελέτες των Duan και συν. (2012), Brooks και συν. (2014), Li και συν. (2016), Lu και συν. (2016), Hao και συν. (2013), Torkildson και συν. (2014), και Chen και συν. (2015).

Τέλος, η υλοποιηθείσα εφαρμογή χρησιμοποιεί μεγάλη ποικιλία τεχνικών, που κατηγοριοποιούνται με βάση τα χαρακτηριστικά οπτικοποίησης (visualization aspects) και τον τύπο των χρησιμοποιούμενων γραφικών, όπως αυτές που έχουν παρουσιαστεί στην Ενότητα 2.5.5. Ειδικότερα, με βάση τις ιδιότητες των γραφικών (visual variables), χρησιμοποιείται η πολύ συνήθης πρακτική απεικόνισης της πολικότητας συναισθημάτων με χρωματική διαβάθμιση, όπως έχει γίνει και στις ερευνητικές εργασίες των Wensel και Sood (2008), Tsirakis και συν. (2016), Wu και συν. (2012; 2013), Calderon και συν. (2014), Chen και συν. (2014), Hoque και Carenini (2014), Hoque και Carenini (2016), Yu και συν. (2016), Kang και Ren (2011), Zhao και συν. (2014a), El-Assady και συν. (2016), Almutairi (2013) και Wang και συν. (2015). Επιπρόσθετα, το μέγεθος ή το εύρος των σχηματικών αναπαραστάσεων χρησιμοποιούνται ευρύτατα ως απεικονιστικές τεχνικές για να γίνει εύκολα αντιληπτή η κατανόηση του μεγέθους ή της ποσότητας των απεικονιζόμενων μεταβλητών, κυρίως μέσα από ραβδογράμματα ή άλλα αντίστοιχα γραφήματα. Πληθώρα τέτοιων τεχνικών συναντώνται στη σχετική βιβλιογραφία, όπως παρουσιάζεται αναλυτικά στην Ενότητα 2.5.5 (Alper et al., 2011; Mohammad et al., 2016; Marcus et al., 2011; Das et al., 2012; Liu et al., 2013; Kranjc et al., 2015; Kempter et al., 2014; Munezero et al., 2015; Gamon et al., 2008; Wecker

et al., 2014). Επίσης, η οπτικοποίηση με χρήση γλύφων, που χρησιμοποιούνται στην παρούσα υλοποίηση, συναντώνται ευρύτατα και στη σχετική βιβλιογραφία (Liu et al., 2003; Gobron et al., 2010; Lee et al., 2010; Cresci et al., 2014; Li & Ren, 2009; Almutairi, 2013; Kempter et al., 2014; Munezero et al., 2015).

4.2 Σύνοψη και συμπεράσματα

Λόγω της έκρηξης του όγκου αλλά και του βαθμού διακίνησης των δεδομένων, που σήμερα υπάρχουν διαθέσιμα από πολλαπλές πηγές και σε πολλαπλές μορφές, και τα οποία παρέχουν τεράστιες δυνατότητες κατανόησης των συναισθημάτων, αρκεί κανείς να επιτύχει την αποτελεσματική επεξεργασία τους με μαζικό ή επιλεκτικό τρόπο, φιλτράροντας τον όποιο θόρυβο τα περιβάλλει. Ένα από τα σύγχρονα ερευνητικά προβλήματα που έχει επομένως αναδυθεί τα τελευταία χρόνια είναι η προσπάθεια αυτοματοποιημένης ανάλυση συναισθημάτων, όρος που συναντάται και ως εξόρυξη γνώμης, και έχει ως στόχο την ανίχνευση και ανάλυση κειμένων και ταξινόμησή των γραφόμενων ανάλογα με τα συναισθήματα που εμπεριέχουν (Kucher et al., 2018, Liu, 2009; Pang & Lee, 2008).

Ως ανάλυση συναισθημάτων ορίζεται συνήθως η διαδικασία κατηγοριοποίησης τμημάτων κειμένου, που εκτείνονται από μεμονωμένες λέξεις και φράσεις ή προτάσεις έως πλήρη έγγραφα, σε ένα μικρό αριθμό κατηγοριών που αντιπροσωπεύουν διαφορετικά είδη συναισθημάτων. Ο όρος συναίσθημα, στη συγκεκριμένη επιστημονική περιοχή, αποτελεί συνήθως συνώνυμο όρων όπως συγκίνηση, επίδραση, στάση κ.λπ. Στην απλούστερη διατύπωσή της, η ανάλυση συναισθημάτων θεωρείται δυαδικό πρόβλημα, κατά την επίλυση του οποίου εντοπίζονται και διαχωρίζονται συναισθηματικά τα θετικά από τα αρνητικά συναισθήματα, διαδικασία γνωστή και ως ανίχνευση πολικότητας (Turney, 2002). Μεταγενέστερες ερευνητικές προσπάθειες υιοθέτησαν μια προσέγγιση, κατά την οποία στην ανάλυση συναισθημάτων πραγματοποιείται κατηγοριοποίησή τους σε τρεις κατηγορίες, αρνητικά, ουδέτερα και θετικά, ενώ υπάρχει και βιβλιογραφία που χρησιμοποιεί κλιμακωτή διαβάθμιση σε περισσότερες από τρεις κατηγορίες, για παράδειγμα ασθενώς έως έντονα αρνητική, ουδέτερη, ασθενώς ως έντονα θετική (Socher et al., 2013; Socher et al., 2011; He, 2012; Amigó et al., 2013).

Η οπτικοποίηση συναισθημάτων αναφέρεται στους τρόπους και τις τεχνικές, μέσα από τις οποίες δύνανται να απεικονιστούν και να αναλυθούν οι πληροφορίες ανάλυσης συναισθημάτων, που προέκυψαν από τη συλλογή και επεξεργασία δεδομένων κειμένου. Η οπτικοποίηση συναισθημάτων αποτελεί υποσύνολο της ευρύτερης ερευνητικής περιοχής, που είναι γνωστή ως οπτικοποίηση κειμένου. Η οπτικοποίηση των αναλύσεων συναισθημάτων και των στοιχείων που προκύπτουν από την επεξεργασία των σχετικών κειμένων αποτελεί ένα σύγχρονο πεδίο ερευνητικής δραστηριότητας εδώ και περισσότερο από μία δεκαετία. Τα αποτελέσματα αναλύσεων συναισθημάτων οπτικοποιούνται πλέον με πολλούς και διαφορετικούς τρόπους, από παραδοσιακά γραφήματα πίτας ή ραβδογράμματα μέχρι εξειδικευμένα γραφήματα οπτικοποίησης, που περιλαμβάνουν νέους τρόπους αναπαράστασης κατάλληλους για την ανάλυση πολύπλοκων και πολυδιάστατων συνόλων δεδομένων, που συνδυάζουν μεταξύ άλλων τις διαστάσεις χρόνου και χώρου και μπορούν την ίδια στιγμή να περιγράψουν αλληλοσυσχετίσεις (Kucher & Kerren, 2015; Kucher et al., 2018).

Η ερευνητική περιοχή της οπτικοποίησης συναισθημάτων εμπεριέχει διαφορετικά θεματικά πεδία, όπως είναι η παρακολούθηση της κοινής γνώμης στα μέσα κοινωνικής δικτύωσης, η ψηφιακή ανάλυση βιβλιογραφίας ανθρωπιστικών επιστημών, η έρευνα ανάλυσης συναισθημάτων, η γλωσσολογία και η επεξεργασία φυσικής γλώσσας (Natural Language Processing). Κατά τις πρώτες φάσεις εξέλιξης του συγκεκριμένου ερευνητικού πεδίου η οπτικοποίηση συναισθημάτων συσχετίστηκε με την εξόρυξη δεδομένων (data mining) και την επεξεργασία φυσικής γλώσσας, ενώ οι τεχνικές οπτικοποίησης ήταν κατά κανόνα απλές. Πλέον, οι τεχνικές που σήμερα χρησιμοποιούνται αντανακλούν την πρόοδο που έχει σημειωθεί τα τελευταία χρόνια στο ευρύτερο πεδίο της οπτικοποίησης, είναι πολύ πιο σύνθετες και διαχειρίζονται ευρεία γκάμα ετερογενών δεδομένων (Kucher et al., 2018).

Συμπερασματικά, θα μπορούσε κανείς να πει ότι η παρούσα εργασία πραγματεύεται το θέμα της ανάλυσης συναισθημάτων και ειδικότερα το πώς μπορούν να παρουσιαστούν τα αποτελέσματα μια τέτοιας ανάλυσης, ώστε να μπορούν να γίνουν ευκολότερα κατανοητά και να αξιοποιηθούν με αποτελεσματικότερο τρόπο. Στο πρώτο τμήμα της εργασίας έχει πραγματοποιηθεί ευρεία βιβλιογραφική ανασκόπηση του θεματικού πεδίου ανάλυσης και οπτικοποίησης συναισθημάτων. Η εν λόγω ανασκόπηση περιλαμβάνει την παρουσίαση και επεξήγηση των πλέον χαρακτηριστικών τρόπων με τους οποίους μπορεί να αξιοποιηθεί η ανάλυση συναισθημάτων από επιχειρήσεις και οργανισμούς, ώστε να βελτιστοποιηθεί η εμπειρία που προσφέρεται στο κοινό, τους τελικούς πελάτες ή τους συναλλασσόμενους. Σημαντικές περιοχές στις οποίες βρίσκει εφαρμογή η ανάλυση

συναισθημάτων, είναι η μελέτη του ανταγωνισμού μιας επιχείρησης και ο προσδιορισμός υποστηρικτών και επικριτών επί ενός θέματος, ή μιας άποψης. Επιπρόσθετα, παρουσιάζεται λεπτομερής κατηγοριοποίηση των τεχνικών οπτικοποίησης συναισθημάτων, ομαδοποιημένων σε κατηγορίες ανάλογα με τα χαρακτηριστικά των χρησιμοποιούμενων δεδομένων, τις πηγές προέλευσης και ιδιότητες των δεδομένων, τις εργασίες που εκτελούνται και τις παραμέτρους και μεταβλητές οπτικοποίησης. Τέλος, έχει πραγματοποιηθεί αναλυτική καταγραφή εργαλείων και εφαρμογών (εμπορικών και ανοιχτού κώδικα – open source), τα οποία δύναται να χρησιμοποιηθούν για την οπτικοποίηση αποτελεσμάτων ανάλυσης συναισθημάτων. Τα συγκεκριμένα εργαλεία χρησιμοποιούν πλήθος διαφορετικών τεχνικών ανάλυσης, μεταξύ των οποίων αλγόριθμους στατιστικής ανάλυσης και ομαδοποίησης (statistical analysis and clustering), τεχνητής νοημοσύνης (artificial intelligence) και μηχανικής και σε βάθος μάθησης (machine and deep learning), ενώ υφίσταται επίσης μεγάλο πλήθος εφαρμογών που έχουν αναπτυχθεί πάνω σε ανοιχτού κώδικα προγραμματιστικές πλατφόρμες και εργαλεία (π.χ. Python, R, κ.λπ.).

Στο δεύτερο τμήμα της εργασίας παρουσιάζεται η μελέτης περίπτωσης (Case Study), στο πλαίσιο της οποίας έχει αναπτυχθεί διαδραστική διεπαφή μέσω της οποίας παρουσιάζονται, αναλύονται και αξιολογούνται τα αποτελέσματα ανάλυσης συναισθημάτων. Για το σκοπό αυτό έχει χρησιμοποιηθεί ένα σύνολο δεδομένων (data set), το οποίο περιλαμβάνει κριτικές και αξιολογήσεις πελατών για προϊόντα και υπηρεσίες που παρέχονται από μια εταιρία πώλησης γυναικείων ρούχων. Για την ανάλυση συναισθημάτων και οπτικοποίηση των αποτελεσμάτων έχει χρησιμοποιηθεί η γλώσσα Python, περιλαμβανομένων και δημοφιλών βιβλιοθηκών γραφικής απεικόνισης.

Η διεπαφή είναι προσπελάσιμη και απεικονίζει τα γραφικά μέσω περιηγητή ιστοσελίδων (web browser). Για το κείμενο που αναλύεται, απεικονίζονται πολλαπλά γραφικά και πληροφορίες, τα οποία συνδυάζονται και ομαδοποιούνται με λογικό τρόπο ανά σελίδα. Η εφαρμογή είναι πολλαπλών σελίδων (multi page application), οι οποίες έχουν μια λογική αλληλουχία. Οι λογικές ενότητες/ σελίδες της web-εφαρμογής, που ομαδοποιούν τις πληροφορίες και τα γραφικά τα οποία απεικονίζονται μαζί, είναι οι ακόλουθες:

- *Ενότητα A – Data Sample Basics*: Εισαγωγική σελίδα, η οποία περιλαμβάνει βασικές πληροφορίες για το σύνολο δεδομένων (data set).
- *Ενότητα B - Visualizations*: Ομάδα ραβδογραμμάτων (bar charts), που οπτικοποιούν τη στατιστική ανάλυση των κριτικών και την κατανομή τους βάσει διαφορετικών κριτηρίων.

- *Ενότητα Γ – Box Plots*: Γραφήματα τύπου box plot, που απεικονίζουν με ιδιαίτερα παραστατικό και διαδραστικό τρόπο τα συναισθήματα βάσει διαφορετικών κριτηρίων.
- *Ενότητα Δ – Latent Analysis*: Γραφική απεικόνιση λανθάνουσας ανάλυσης (latent analysis), που παρουσιάζει τα έξι (6) δημοφιλέστερα θέματα συζήτησης (topics) μέσα από δύο διαφορετικούς τύπους γραφημάτων.

4.3 Όρια και περιορισμοί της έρευνας – Μελλοντικές επεκτάσεις

Η παρούσα εργασία απευθύνεται στους ερευνητές οπτικοποίησης συναισθημάτων, καθώς συγκεντρώνει και παρουσιάζει μια ευρεία βιβλιογραφική ανάλυση της συγκεκριμένης περιοχής, χρησιμοποιώντας ως βάση δημοσιεύσεις, που είναι διάσπαρτες σε μεγάλο αριθμό επιστημονικών υποπεριοχών. Επιπρόσθετα, μπορεί να είναι χρήσιμη και για ερευνητές άλλων επιστημονικών πεδίων αλλά και για επαγγελματίες, που ενδιαφέρονται για μεθόδους οπτικής ανάλυσης δεδομένων συναισθημάτων, αφού πρόκειται για μια αγορά που αποτιμάται περί τα 22.65 δισεκατομμύρια δολάρια.

Από θεωρητική σκοπιά, η παρούσα εργασία έχει μεν διερευνήσει σε σημαντικό βαθμό τη σχετική βιβλιογραφία, όμως θα πρέπει να είναι ξεκάθαρο ότι δεν εστιάζει στο θεωρητικό υπόβαθρο αυτής καθεαυτής της ανάλυσης συναισθημάτων, αλλά στη σχετική έρευνα και εφαρμογές που υπάρχουν σχετικά με την οπτικοποίηση της ανάλυσης συναισθημάτων. Στη συγκεκριμένη περιοχή, όχι μόνο έχει διερευνηθεί η σχετική βιβλιογραφία, αλλά έχει πραγματοποιηθεί και ανασκόπηση των σχετικών εργαλείων και του χρησιμοποιούμενου λογισμικού. Επιπρόσθετα, η παρούσα εργασία υπερέβη τα όρια της βιβλιογραφικής ανασκόπησης και συμπεριέλαβε και σχετική μελέτη περίπτωσης, ώστε να καταδείξει ενδεικτικούς πρακτικούς τρόπους υλοποίησης εφαρμογών οπτικοποίησης ανάλυσης συναισθημάτων. Συνολικά λοιπόν, μπορεί κανείς εύκολα να κατανοήσει ότι τα όρια και οι περιορισμοί που τίθενται στο πλαίσιο της παρούσας μελέτης είναι αυτά που ενδεχομένως θέτουν τα χρησιμοποιηθέντα εργαλεία βιβλιογραφικής αναζήτησης και ανάπτυξης σχετικών εφαρμογών, τα οποία εν προκειμένω είναι ανοιχτού κώδικα (open source).

Ως προς τη μελλοντική επέκταση της παρούσας έρευνας, αυτή θα μπορούσε να κινηθεί προς τρεις βασικές κατευθύνσεις. Η πρώτη κατεύθυνση είναι η περαιτέρω εμβάθυνση στη θεωρητική θεμελίωση της ανάλυσης και οπτικοποίησης συναισθημάτων. Αυτό μπορεί να γίνει με μεγαλύτερη

διερεύνηση, ανάπτυξη και τεκμηρίωση του θεωρητικού υπόβαθρου μέσα από τη σχετική βιβλιογραφία. Η δεύτερη κατεύθυνση είναι η περαιτέρω επέκταση της παρούσας υλοποίησης σε επόμενες εκδόσεις, οι οποίες θα μπορούσαν να συμπεριλάβουν περισσότερες και αναλυτικότερες διαδραστικές και δυναμικές οπτικοποιήσεις και ακόμη μεγαλύτερο βαθμό ανάλυσης των αποτελεσμάτων. Τέλος, μία ακόμη κατεύθυνση επέκτασης της παρούσας υλοποίησης είναι η χρήση επιπλέον γραφικών βιβλιοθηκών της Python, ή ακόμη και η υλοποίηση με χρήση κάποιας άλλης πλατφόρμας ή άλλου λογισμικού.

Βιβλιογραφία

- Abbasi, A., & Chen, H. (2007). Categorization and analysis of text in computer mediated communication archives using visualization. *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '07, ACM*, pp. 11-18. <https://doi.org/10.1145/1255175.1255178>
- Almutairi, B. (2013). Visualizing patterns of appraisal in texts and corpora. *Text & Talk*, 33(4-5), 691-723. <https://doi.org/10.1515/text-2013-0031>
- Alper, B., Yang, H., Haber, E., Kandogan, E. (2011). OpinionBlocks: Visualizing consumer reviews. *Proceedings of the IEEE Workshop on Interactive Visual Text Analytics for Decision Making, TextVis '11*. <https://bit.ly/2YQcdwI>
- Amigó, E., De Albornoz, J., Chugur, I., Coruogo, A., Gonzalo, J., Martin, T., Meij, E., De Rijke, M., & Spina, D. (2013). Overview of RepLab 2013: Evaluating online reputation monitoring systems. *Information Access Evaluation. Multilinguality, ultimodality, and Visualization. Springer, 2013*, pp. 333-352. https://doi.org/10.1007/978-3-642-40802-1_31
- Annett, M., & Kondrak, G. (2008). A comparison of sentiment analysis techniques: Polarizing movie blogs. *Advances in Artificial Intelligence, 5032*, 25-35. https://doi.org/10.1007/978-3-540-68825-9_3
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the Seventh Conference on International Language Resources and Evaluation, vol. 10 of LREC '10, European Language Resources Association*, pp. 2200-2204. <https://bit.ly/3d9VF7a>
- Bembenik, R., & Andruszkiewicz, P. (2016). Towards automatic argument extraction and visualization in a deliberative model of online consultations for local governments. In *Proceedings of the 20th East European Conference on Advances in Databases and Information Systems (ADBIS 2016)*, Pokorny J., Ivanovic M., Thalheim B., Saloun P., (Eds.), vol. 9809 of Lecture Notes in Computer Science. Springer International Publishing, pp. 74-86. https://doi.org/10.1007/978-3-319-44039-2_6
- Boumaiza, A. (2015). A survey on sentiment analysis and visualization. *Journal of Emerging Technologies in Web Intelligence*, 7(1), 35-43. <https://bit.ly/2N1SwLX>
- Bras-Oveanu, A., Hubmann-Haidvogel, A., & Scharl, A. (2012). Interactive visualization of emerging topics in multiple social media streams. *Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '12, ACM*, pp. 530-533. <https://doi.org/10.1145/2254556.2254655>
- Brath, R., & Banissi, E. (2016). Typographic sets: Labeled set elements with font attributes. *Proceedings of the International Workshop on Set Visualization and Reasoning, SetVR 2016*, pp. 29-43. <https://bit.ly/2NOSneY>

- Brew, A., Greene, D., Archambault, D., & Cunningham, P. (2011). Deriving insights from national happiness indices. *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, ICDMW '11*, pp. 53-60. <https://doi.org/10.1109/ICDMW.2011.61>
- Brody, S., & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10, Association for Computational Linguistics*, pp. 804-812. <https://bit.ly/3hrzD2R>
- Brooks, M., Robinson, J., Torkildson, M., Hong, S., & Aragon, C. (2014). Collaborative visual analysis of sentiment in Twitter events. *Proceedings of the 11th International Conference on Cooperative Design, Visualization, and Engineering (CDVE '14), Luo Y., (Ed.), Vol. 8683 of Lecture Notes in Computer Science. Springer International Publishing*, pp. 1-8. <https://bit.ly/3dRO8dn>
- Calderon, N., Arias-Hernandez, R., & Fisher, B. (2014). Studying animation for real-time visual analytics: A design study of social media analytics in emergency management. *Proceedings of the 47th Hawaii International Conference on System Sciences, HICSS '14*, pp. 1364-1373. <https://doi.org/10.1109/HICSS.2014.176>
- Cambria E., Olsher, D., & Rajagopal, D. (2014). SenticNet 3: A common and common-sense knowledge base for cognition driven sentiment analysis. *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI '14, AAAI Press*, pp. 1515-1521. <https://bit.ly/2Y3GJ6b>
- Cao, N., Lin, Y.-R., Sun, X., Lazer, D., Liu, S., & Qu, H. (2012). Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2649-2658. <https://doi.org/10.1109/TVCG.2012.291>
- Cao, N., Lu, L., Lin, Y.-R., Wang, F., Wen, Z. (2015). SocialHelix: visual analysis of sentiment divergence in social media. *Journal of Visualization*, 18(2), 221-235. <https://doi.org/10.1007/s12650-014-0246-x>
- Cao, N., Shi, C., Lin, S., Lu, J., Lin, Y.-R., & Lin, C.-Y. (2016). TargetVue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 280-289. <https://bit.ly/2Y0tH9n>
- Caragea, C., Squicciarini, A., Stehle, S., Neppalli, K., & Tapia, A. (2014). Mapping moods: Geo-mapped sentiment analysis during Hurricane Sandy. *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management, ISCRAM '14*. <https://bit.ly/3eqZl5A>
- Cervantes, O., Gutierrez, F., Gutierrez, E., Castillo, E., Sanchez, J., & Wan, W. (2015). Expression: Visualizing affective content from social streams. *Proceedings of the Latin American Conference on Human Computer Interaction, CLIHC '15, ACM*, pp. 10:1-10:8. <https://doi.org/10.1145/2824893.2824903>

- Chen, C., Ibekwe-Sanjuan, F., Sanjuan, E., & Weaver, C. (2006). Visual analysis of conflicting opinions. *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (2006), VAST '06, IEEE*, pp. 59-66. <https://bit.ly/2YzVPzb>
- Chen, N.-C., Brooks, M., Kocielnik, R., Hong, S., Smith, J., Lin, S., Qu, Z., & Aragon, C. (2017). Lariat: A visual analytics tool for social media researchers to explore Twitter datasets. *Proceedings of the 50th Hawaii International Conference on System Sciences, HICSS '17*, pp. 1881-1890. <https://bit.ly/3dKJpdj>
- Chen, N.-C., Feldman, L., Kroll, J., & Aragon, C. (2014). Emoticons and linguistic alignment: How visual analytics can elicit storytelling. *Poster Abstracts of IEEE VAST, VAST '14*, pp. 237-238. <https://bit.ly/2MzMzz4>
- Chen, Y.-S., Chen, L.-H., Takama, Y. (2015). Proposal of LDA based sentiment visualization of hotel reviews. *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop, ICDMW 2015*, pp. 687-693. <https://doi.org/10.1109/ICDMW.2015.72>
- Cresci, S., Derrivco, A., Gazze, D., Lo Duca, A., Marchetti, A., & Tesconi, M. (2014). Tour-pedia: A web application for sentiment visualization in tourism domain. *Proceedings of the LREC 2014 OpeNER Workshop*, pp. 18-21. <https://bit.ly/2ZscTaP>
- Cui, W., Qu, H., Zhou, H., Zhang, W., & Skiena, S. (2012). Watch the story unfold with TextWheel: Visualization of large-scale news streams. *ACM Transactions on Intelligent Systems and Technology*, 3(2), 20:1-20:17. <https://bit.ly/2V9FwZ3>
- Dai, X., & Prout, R. (2016). Unlocking Super Bowl insights: Weighted word embeddings for Twitter sentiment classification. *Proceedings of the 3rd Multidisciplinary International Social Networks Conference, SocialInformatics, Data Science, MISNC, SI, DS 2016, ACM*, pp. 12: 1-12:6. <https://doi.org/10.1145/2955129.2955148>
- Das, A., Bandyopadhyay, S., & Gamback, B. (2012). Sentiment analysis: What is the end user's requirement? *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, WIMS '12, ACM*, pp. 35:1-35:10. <https://bit.ly/30Iya2h>
- Dehiya, V., & Mueller, K. (2016). Analyzing Hillary Clinton's emails. *Poster Abstracts of IEEE VIS, InfoVis '16*. <https://bit.ly/3ge0Zlq>
- Di Caro, L., & Grella, M. (2013). Sentiment analysis via dependency parsing. *Computer Standards & Interfaces*, 35(5), 442-453. <https://doi.org/10.1016/j.csi.2012.10.005>
- Diakopoulos, N., Naaman, M., & Kivran-Swaine, F. (2010). Diamonds in the rough: Social media visual analytics for journalistic inquiry. *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, VAST '10, IEEE*, pp. 115-122. <https://bit.ly/3e4BCrX>
- Diakopoulos, N., Zhang, A., Elgesem, D., & Salway, A. (2014). Identifying and analyzing moral evaluation frames in climate change blog discourse. *Proceedings of the International Conference on Weblogs and Social Media, ICWSM '14*, pp. 583-586. <https://bit.ly/3ikXX76>

- Duan, D., Qian, W., Pan, S., Shi, L., & Lin, C. (2012). VISA: A visual sentiment analysis system. *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction, VINCI '12, ACM*, pp. 22-28. <https://bit.ly/2NmMjdG>
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3/4), 169-200. <https://bit.ly/3hwgZqP>
- El-Assady, M., Gold, V., Acevedo, C., Collins, C., & Keim., D. (2016). ConToVi: Multi-party conversation exploration using topicspace views. *Computer Graphics Forum*, 35(3), 431-440. <https://doi.org/10.1111/cgf.12919>
- Farina, J., Mazuran, M., & Quintarelli, E. (2014). Extraction, Sentiment Analysis and Visualization of Massive Public Messages. In: Catania, B. et al. (Eds.) *New Trends in Databases and Information Systems. Advances in Intelligent Systems and Computing*, Vol 241. Springer, Cham.
- Fu, S., Zhao, J., Cui, W., & Qu, H. (2017). Visual analysis of MOOC forums with iForum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 201-210. <https://bit.ly/3eQqnU9>
- Fukuhara, T., Nakagawa, H., & Nishida, T. (2007). Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. *Proceedings of The 2nd International Conference on Weblogs and Social Media, ICWSM '07*. <https://bit.ly/314xwfb>
- Gali, G., Oliver, S., Chevalier, F., & Diamond, S. (2012). Visualizing sentiments in business-customer relations with metaphors. *CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA '12, ACM*, pp. 1493-1498. <https://bit.ly/2BA19L1>
- Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. *Proceedings of 6th International Symposium on Intelligent Data Analysis (IDA 2005)*, Famili A. F., Kok J. N., Peña J. M., Siebes A., Feelders A., (Eds.), vol. 3646 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 121-132. https://doi.org/10.1007/11552253_12
- Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, M., & Konig, A. (2008). BLEWS: Using blogs to provide context for news articles. *Proceedings of the International Conference on Weblogs and Social Media, ICWSM '08*, pp. 60-67. <https://bit.ly/30IqTzB>
- Gobron, S., Ahn, J., Paltoglou, G., Thelwall, M., & Thalmann, D. (2010). From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text. *The Visual Computer*, 26(6), 505-519. <https://bit.ly/2B7fr5K>
- Gold, V., Rohrsantz, C., & El-Assady, M. (2015). Exploratory text analysis using lexical episode plots. In Short Papers of the EG/VGTC Conference on Visualization, Bertini E., Kennedy J., Puppo E., (Eds.), EuroVis '15, The Eurographics Association. <https://doi.org/10.2312/eurovisshort.20151130>
- Gorg, C., Liu, Z., Kihm, J., Choo, J., Park, H., & Stasko, J. (2013). Combining computational analyses and interactive visualization for document exploration and sensemaking in Jigsaw.

- IEEE Transactions on Visualization and Computer Graphics*, 19(10), 1646-1663.
<https://bit.ly/3dTixbc>
- Gregory, M., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., & Turner, A. (2006). User-directed sentiment analysis: Visualizing the affective content of documents. *Proceedings of the Workshop on Sentiment and Subjectivity in Text, SST '06, Association for Computational Linguistics*, pp. 23-30. <https://bit.ly/38gtSkq>
- Guzman, E. (2013). Visualizing emotions in software development projects. *Proceedings of the 2013 First IEEE Working Conference on Software Visualization, VISSOFT '13, IEEE*, pp. 1-4. <https://doi.org/10.1109/VISSOFT.2013.6650529>
- Guzman, E., Bhuvanagiri, P., & Bruegge, B. (2014). FAVe: Visualizing user feedback for software evolution. *Proceedings of the 2014 Second IEEE Working Conference on Software Visualization, VISSOFT '14, IEEE*, pp. 167-171. <https://doi.org/10.1109/VISSOFT.2014.33>
- Ha, H., Kim, G.-N., Hwang, W., Choi, H., & Lee, K. (2014). CosMovis: Analyzing semantic network of sentiment words in movie reviews. *Poster Abstracts of IEEE LDAH, LDAH '14*, pp. 113-114. <https://bit.ly/2BvrNVV>
- Hao, M. C., Rohrdantz, C., Janetzko, H., Keim, D., Dayal, U., Haug, L.-E., Hsu, M., & Stoffel, F. (2013). Visual sentiment analysis of customer feedback streams using geo-temporal term associations. *Information Visualization*, 12(3-4), 273-290.
<https://doi.org/10.1177/1473871613481691>
- He, Y. (2012). A Bayesian modeling approach to multi-dimensional sentiment distributions prediction. *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining (2012), WISDOM '12, ACM*, pp. 1:1-1:8.
<https://doi.org/10.1145/2346676.2346677>
- Hennig, P., Berger, P., Brehm, M., Grasnack, B., Herdt, J., & Meinel, C. (2015). Hot spot detection - an interactive cluster heat map for sentiment analysis. *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015*.
<https://doi.org/10.1109/DSAA.2015.7344885>
- Hennig, P., Berger, P., Meinel, C., Pirl, L., & Schulze, L. (2014). Exploring emotions over time within the blogosphere. *Proceedings of the 2014 International Conference on Data Science and Advanced Analytics, DSAA 2014*, pp. 587-592.
<https://doi.org/10.1109/DSAA.2014.7058131>
- Hoerber, O., Hoerber, L., Meseery, M., Odoh, K., & Gopi, R. (2016). Visual Twitter analytics (Vista): Temporally changing sentiment and the discovery of emergent themes within sport event tweets. *Online Information Review*, 40(1), 25-41. <https://doi.org/10.1108/OIR-02-2015-0067>
- Hogenboom, A., Heerschop, B., Frasinca, F., Kaymak, U., & De Jong, F. (2014). Multi-lingual support for lexicon-based sentiment analysis guided by semantics. *Decision Support Systems*, 62, 43-53. <https://doi.org/10.1016/j.dss.2014.03.004>

- Hoque, E., & Carenini, G. (2014). ConVis: A visual text analytic system for exploring blog conversations. *Computer Graphics Forum*, 33(3), 221-230. <https://bit.ly/3hUtoEX>
- Hoque, E., Carenini, G. (2016). MultiConVis: A visual text analytics system for exploring a collection of online conversations. *Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16, ACM*, pp. 96-107. <https://bit.ly/3159Ac4>
- Jo, Y., & Oh A. (2011). Aspect and sentiment unification model for online review analysis. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, ACM*, pp. 815-824. <https://bit.ly/30I7AX2>
- Kang, X., & Ren, F. (2011). Sampling latent emotions and topics in a hierarchical bayesian network. *Proceedings of the 2011 7th International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE '11, IEEE*, pp. 37-42. <https://doi.org/10.1109/NLPKE.2011.6138166>
- Kempton, R., Sintsova, V., Musat, C., & Pu, P. (2014). Emotion-Watch: Visualizing fine-grained emotions in event-related tweets. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, ICWSM '14*, pp. 236-245. <https://bit.ly/2NKbmYh>
- Kim, K., & Lee, J. (2014). Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition*, 47(2), 758-768. <https://doi.org/10.1016/j.patcog.2013.07.022>
- Kranjc, J., Smailovic, J., Podpecan, V., Grcar, M., Znidarsic, M., & Lavrac, N. (2015). Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the ClowdFlows platform. *Information Processing & Management*, 51(2), 187-203. <https://doi.org/10.1016/j.ipm.2014.04.001>
- Krcadinac, U., Jovanovic, J., Devedzic, V., & Pasquier, P. (2016). Textual affect communication and evocation using abstract generative visuals. *IEEE Transactions on Human-Machine Systems*, 46(3), 370-379. <https://doi.org/10.1109/THMS.2015.2504081>
- Krstajic, M., Rohrdantz, C., Hund, M., & Weiler, A. (2012). Getting there first: Real-time detection of real-world incidents on Twitter. *Proceedings of the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media", TextVis '12*. <https://bit.ly/38gyuHa>
- Kuang, C., Tang, J., Liu, Z., & Sun, M. (2014). ImgWordle: Image and text visualization for events in microblogging services. *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces, AVI '14, ACM*, pp. 371-372. <https://doi.org/10.1145/2598153.2600041>
- Kucher, K., & Kerren, A. (2015). Text visualization techniques: Taxonomy, visual survey, and community insights. *Proceedings of the 8th IEEE Pacific Visualization Symposium, PacificVis '15, IEEE*, pp. 117-121. <https://bit.ly/2BegW1V>
- Kucher, K., Paradis, C., & Kerren, A. (2018). The State of the Art in Sentiment Visualization. *Computer Graphics Forum*, 37(1), 71-96. <https://doi.org/10.1111/cgf.13217>

- Kucher, K., Schamp-Bjerede, T., Kerren, A., Paradis, C., & Sahlgren, M. (2016). Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization*, 15(2), 93-116. <https://doi.org/10.1177/14738716155575079>
- Kuksenok, K., Brooks, M., Robinson, J., Perry, D., Torkildson, M., & Aragon, C. (2012). Automating large-scale annotation for analysis of social media content. *Proceedings of the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media"*, TextVis '12. <https://bit.ly/2NLrBV0>
- Kumamoto, T., Wada, H., & Suzuki, T. (2014). Visualizing temporal changes in impressions from tweets. *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services, iiWAS '14, ACM*, pp. 116-125. <https://doi.org/10.1145/2684200.2684279>
- Kwon, B., Kim, S.-H., Lee, S., Choo, J., Huh, J., & Yi, J. (2016). VisOHC: Designing visual analytics for online health communities. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 71-80. <https://doi.org/10.1109/TVCG.2015.2467555>
- Lee, H., Ferguson, P., Ohare, N., Gurrin, C., & Smeaton, A. (2010). Integrating interactivity into visualising sentiment analysis of blogs. *Proceedings of the First International Workshop on Intelligent Visual Interfaces for Text Analysis, IVITA '10, ACM*, pp. 17-20. <https://doi.org/10.1145/2002353.2002360>
- Li, H., & Ren, F. (2009). The study on text emotional orientation based on a three-dimensional emotion space model. *Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE '09, IEEE*. <https://doi.org/10.1109/NLPKE.2009.5313815>
- Li, Q., Wu, Y., Wang, S., Lin, M., Feng, X., & Wang, H. (2016). VisTravel: visualizing tourism network opinion from the user generated content. *Journal of Visualization*, 19(3), 489-502. <https://doi.org/10.1007/s12650-015-0330-x>
- Liu, B. (2009). Sentiment Analysis and Subjectivity. In: N. Indurkha & F.J. Damerau (Eds.) *Handbook of Natural Language Processing* (2nd Ed.), Marcel Dekker Inc., New York. <https://bit.ly/3dNzp3M>
- Liu, B., Hu, M., & Cheng, J. (2005). Opinion Observer: Analyzing and comparing opinions on the Web. *Proceedings of the 14th International Conference on WorldWideWeb, WWW'05, ACM*, pp. 342-351. <https://bit.ly/2UNr2hc>
- Liu, H., Selker, T., & Lieberman, H. (2003). Visualizing the affective structure of a text document. *CHI '03 Extended Abstracts on Human Factors in Computing Systems, CHI EA '03, ACM*, pp. 740-741. <https://bit.ly/2V7ATyu>
- Liu, S., Zhu, W., Xu, N., Li, F., Cheng, X.-Q., Liu, Y., & Wang, Y. (2013). Co-training and visualizing sentiment evolution for tweet events. *Proceedings of the 22nd International Conference on World Wide Web, WWW'13 Companion, ACM*, pp. 105-106. <https://doi.org/10.1145/2487788.2487836>

- Liu, X., Xu, A., Gou, L., Liu, H., Akkiraju, R., & Shen, H.-W. (2016). SocialBrands: Visual analysis of public perceptions of brands on social media. *Proceedings of the IEEE Conference on Visual Analytics Science and Technology, VAST '16*, pp. 71-80. <https://doi.org/10.1109/VAST.2016.7883513>
- Lu, Y., Hu, X., Wang, F., Kumar, S., Liu, H., & Maciejewski, R. (2015). Visualizing social media sentiment in disaster scenarios. *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion, ACM*, pp. 1211-1215. <https://doi.org/10.1145/2740908.2741720>
- Lu, Y., Kruger, R., Thom, D., Wang, F., Koch, S., Ertl, T., & Maciejewski, R. (2014). Integrating predictive analytics and social media. *Proceedings of the IEEE Conference on Visual Analytics Science and Technology, VAST '14, IEEE*, pp. 193-202. <https://bit.ly/37SkzXH>
- Lu, Y., Steptoe, M., Burke, S., Wang, H., Tsai, J.-Y., Davulcu, H., Montgomery, D., Corman, S. R., & Maciejewski, R. (2016). Exploring evolving media discourse through event cueing. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 220-229. <https://doi.org/10.1109/TVCG.2015.2467991>
- Mahmud, J., Fei, G., Xu, A., Pal, A., & Zhou, M. (2016). Predicting attitude and actions of Twitter users. *Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16, ACM*, pp. 2-6. <https://doi.org/10.1145/2856767.2856800>
- Makki, R., Brooks, S., & Milios, E. (2014). Context-specific sentiment lexicon expansion via minimal user interaction. *Proceedings of the International Conference on Information Visualization Theory and Applications, IVAPP '14, SciTePress*, pp. 178-186. <https://bit.ly/2VCbZHM>
- Mäntylä, M., Graziotin, D., & Kuuttila, M. (2018). The evolution of sentiment analysis - A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16-32, <https://doi.org/10.1016/j.cosrev.2017.10.002>
- Marcus A., Bernstein, M., Badar, O., Karger, D., Madden, S., & Miller, R. (2011). TwitInfo: Aggregating and visualizing microblogs for event exploration. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, ACM*, pp. 227-236. <https://bit.ly/37wwzxS>
- Miao, Q., Li, Q., & Dai, R. (2009). AMAZING: A sentiment mining and retrieval system. *Expert Systems with Applications*, 36(3/2), 7192-7198. <https://doi.org/10.1016/j.eswa.2008.09.035>
- Mishne, G., & De Rijke, M. (2006). MoodViews: Tools for blog mood analysis. *Proceedings of the AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pp. 153-154. <https://bit.ly/3e8xDKU>
- Mohammad, S. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4), 730-741. <https://doi.org/10.1016/j.dss.2012.05.030>

- Mohammad, S. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion Measurement*, Meiselman H. L., (Ed.). Woodhead Publishing, pp. 201–237. <https://doi.org/10.1016/B978-0-08-100508-8.00009-6>
- Mohammad, S. M., Kiritchenko, S., Sobhani, P., Zhu, X., Cherry, C. (2016). SemEval-2016 task 6: Detecting stance in tweets. *Proceedings of the International Workshop on Semantic Evaluation, SemEval '16*. <https://bit.ly/3hQjyE1>
- Munezero, M., Montero, C., Mozgovoy, M., & Sutinen, E. (2015). EmoTwitter-a fine-grained visualization system for identifying enduring sentiments in tweets. *Proceedings of the 16th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing '15)*, Gelbukh A., (Ed.), Vol. 9042 of Lecture Notes in Computer Science. Springer International Publishing, pp. 78-91. https://doi.org/10.1007/978-3-319-18117-2_6
- Neviarouskaya, A., Aono, M., Prendinger, H., & Ishizuka, M. (2014). Intelligent interface for textual attitude analysis. *ACM Transactions on Intelligent Systems and Technology*, 5(3), 48:1-48:20. <https://doi.org/10.1145/2535912>
- Norambuena, B., Lettura, E., & Villegas, C. (2019). Sentiment analysis and opinion mining applied to scientific paper reviews. *Intelligent Data Analysis*, 23(1), 191-214. <https://doi.org/10.3233/IDA-173807>
- Oelke, D., Hao, M., Rohrdantz, C., Keim, D., Dayal, U., Haug, L.-E., & Janetzko, H. (2009). Visual opinion analysis of customer feedback data. *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (2009), VAST '09, IEEE*, pp. 187-194. <https://bit.ly/3hxcBaK>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1-135. <http://dx.doi.org/10.1561/1500000011>
- Pino, C., Kavasidis, I., & Spampinato, C. (2016). GeoSentiment: A tool for analyzing geographically distributed event-related sentiments. *Proceedings of the 13th IEEE Annual Consumer Communications Networking Conference, CCNC 2016*, pp. 270-271. <https://bit.ly/2Zkl0WM>
- Prasojo, R., Darari, F., Kacimi, M. (2015). ORCAESTRA: Organizing news comments using aspect, entity and sentiment extraction. *Poster Abstracts of IEEE VIS, InfoVis '15*. <https://bit.ly/3159WPW>
- Pupi, S., Di Pietro, G., & Aliprandi, C. (2014). Ent-it-up. *HCI International 2014 - Posters' Extended Abstracts*, Stephanidis C., (Ed.), Vol. 435 (2), *Communications in Computer and Information Science*. Springer International Publishing, pp. 3-8. <https://bit.ly/2NJuc1w>
- Qiu, R., Ravi, R., & Qiu, L. (2015). Aggregating and visualizing public opinions and sentiment trends on the US higher education. *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services, iiWAS '15, ACM*, pp. 33:1-33:5. <https://doi.org/10.1145/2837185.2837261>

- Research & Markets (2020). *Emotion detection and recognition market—Global forecast to 2020*. [Online]. Ανάκτηση 15 Ιουνίου 2020, από: <https://bit.ly/2YBmATZ>
- Rohrdantz, C., Hao, M., Dayal, U., Haug, L.-E., & Keim, D. (2012). Feature-based visual sentiment analysis of text document streams. *ACM Transactions on Intelligent Systems and Technology*, 3(2), 26:1-26:25. <https://bit.ly/2BhWBjL>
- Ruppert, T., Bannach, A., Bernard, J., Lucke-Tieke, H., Ulmer, A., & Kohlhammer, J. (2016). Supporting collaborative political decision making: An interactive policy process visualization system. *Proceedings of the 9th International Symposium on Visual Information Communication and Interaction, VINCI '16*, ACM, pp. 104-111. <https://doi.org/10.1145/2968220.2968223>
- Russell, J., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3), 273-294. [https://doi.org/10.1016/0092-6566\(77\)90037-X](https://doi.org/10.1016/0092-6566(77)90037-X)
- Schäfer, U., & Spurk, C. (2010). TAKE Scientist's Workbench: Semantic search and citation-based visual navigation in scholar papers. *Proceedings of the IEEE Fourth International Conference on Semantic Computing, ISCS '10*, pp. 317-324. <https://doi.org/10.1109/ICSC.2010.40>
- Scharl, A., Hubmann-Haidvogel, A., Jones, A., Fischl, D., Kamolov, R., Weichselbraun, A., & Rafelsberger, W. (2016). Analyzing the public discourse on works of fiction—detection and visualization of emotion in online coverage about HBO's Game of Thrones. *Information Processing & Management*, 52(1), 129-138. <https://doi.org/10.1016/j.ipm.2015.02.003>
- Scharl, A., Hubmann-Haidvogel, A., Weichselbraun, A., Wohlgenannt, G., Lang, H.-P., & Sabou, M. (2012). Extraction and interactive exploration of knowledge from aggregated news and social media content. *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, EICS '12*, ACM, pp. 163-168. <https://doi.org/10.1145/2305484.2305511>
- Shamim, A., Balakrishnan, V., & Tahir, M. (2015). Evaluation of opinion visualization techniques. *Information Visualization*, 14(4), 339-358. <https://doi.org/10.1177/1473871614550537>
- Sijtsma, B., Qvarfordt, P., & Chen, F. (2016). Tweetviz: Visualizing tweets for business intelligence. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '16*, ACM, pp. 1153-1156. <https://doi.org/10.1145/2911451.2911470>
- Small, H. (2011). Interpreting maps of science using citation context sentiments: a preliminary investigation. *Scientometrics*, 87(2), 373-388. <https://doi.org/10.1007/s11192-011-0349-2>
- Socher, R., Pennington, J., Huang, E., Ng, A., & Manning, C. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11*, Association for Computational Linguistics, pp. 151-161. <https://bit.ly/37unaXy>

- Socher, R., Perelygin A., Wu, J., Chuang, J., Manning, C., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP '13, Association for Computational Linguistics*, pp. 1631-1642. <https://stanford.io/2Y2qTIZ>
- Srapparava, C., & Valitutti, A. (2004). WordNet Affect: an affective extension of WordNet. *Proceedings of the 4th International Conference on Language Resources and Evaluation, vol. 4 of LREC '04, European Language Resources Association*, pp. 1083-1086. <https://bit.ly/2Y5Ekbb>
- Steed, C., Beaver, J., Bogen, P., Drouhard, M., & Pyle, J. (2015). Text stream trend analysis using multiscale visual analytics with applications to social media systems. *Proceedings of the IUI Workshop on Visual Text Analytics, TextVis '15*. <https://bit.ly/3fQ3q3w>
- Subasic, P., & Huettner, A. (2001). Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems*, 9(4), 483-496. <https://bit.ly/30JmvjS>
- Sung, C.-Y., Huang, X.-Y., Shen, Y., Cherng, F.-Y., Lin, W.-C., & Wang, H.-C. (2016). ToPIN: A visual analysis tool for time-anchored comments in online educational videos. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16, ACM*, pp. 2185-2191. <https://doi.org/10.1145/2851581.2892327>
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2), 267-307. https://doi.org/10.1162/COLI_a_00049
- Taskaya, T., Ahmad, K. (2003). Bimodal visualisation: a financial trading case study. *Proceedings of Seventh International Conference on Information Visualization, IV '03, IEEE*, pp. 320-326. <https://doi.org/10.1109/IV.2003.1217997>
- Tat, A., & Carpendale, S. (2006). CrystalChat: Visualizing personal chat history. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences, Vol. 3 of HICSS '06, IEEE*. <https://doi.org/10.1109/HICSS.2006.107>
- Tausczik, Y., & Pennebaker, J. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24-54. <https://bit.ly/2YHRavo>
- Torkildson, M., Starbird, K., & Aragon, C. (2014). Analysis and visualization of sentiment and emotion on crisis tweets. *Proceedings of the 11th International Conference on Cooperative Design, Visualization, and Engineering (CDVE '14), Luo Y., (Ed.), Vol. 8683 of Lecture Notes in Computer Science. Springer International Publishing*, pp. 64-67. <https://bit.ly/3dTnoJs>
- Tsirakis, N., Pouloupoulos, V., Tsantilas, P., & Varlamidis, I. (2016). Large scale opinion mining for social, news and blog data. *Journal of Systems and Software*, 127, 237-248. <https://doi.org/10.1016/j.jss.2016.06.012>

- Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, Association for Computational Linguistics*, pp. 417-424. <https://bit.ly/2B6c0fN>
- Wang, C., Xiao, Z., Liu, Y., Xu, Y., Zhou, A., & Zhang, K. (2013). SentiView: Sentiment analysis and visualization for Internet popular topics. *IEEE Transactions on Human-Machine Systems*, 43(6), 620-630. <https://doi.org/10.1109/THMS.2013.2285047>
- Wang, F., Sallaberry, A., Klein, K., Takatsuka, M., & Roche, M. (2015). SentiCompass: Interactive visualization for exploring and comparing the sentiments of time-varying Twitter data. *Proceedings of the 8th IEEE Pacific Visualization Symposium, PacificVis '15*, pp. 129-133. <https://doi.org/10.1109/PACIFICVIS.2015.7156368>
- Wang, J., Dent, K., & North, C. (2013). Fisheyword cloud for temporal sentiment exploration. *CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI EA '13, ACM*, pp. 1767-1772. <https://doi.org/10.1145/2468356.2468673>
- Wang, X., Liu, S., Chen, Y., Peng, T.-Q., Su, J., Yang, J., & Guo, B. (2016). How ideas flow across multiple social groups. *Proceedings of the IEEE Conference on Visual Analytics Science and Technology, VAST '16*, pp. 51-60. <https://doi.org/10.1109/VAST.2016.7883511>
- Wang, Y., Liu, D., Qu, H., Luo, Q., & Ma, X. (2016). A guided tour of literature review: Facilitating academic paper reading with narrative visualization. *Proceedings of the 9th International Symposium on Visual Information Communication and Interaction, VINCI '16, ACM*, pp. 17-24. <https://bit.ly/3diKEjS>
- Wanner, F., Ramm, T., & Keim, D. (2011). ForAVis: Explorative user forum analysis. *Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS '11, ACM*, pp. 14:1-14:10. <https://bit.ly/2UKY69U>
- Wanner, F., Rohrdantz, C., Mansmann, F., Oelke, D., & Keim, D. (2009). Visual sentiment analysis of RSS news feeds featuring the US presidential election in 2008. *Proceedings of the IUI Workshop on Visual Interfaces to the Social and the Semantic Web, VISSW '09*, pp. 1-8. <https://bit.ly/2NnxRCd>
- Wanner, F., Stoffel, A., Jackle, D., Kwon, B., Weiler, A., & Keim, D. A. (2014). State-of-the-art report of visual analysis for event detection in text data streams. *Proceedings of the Eurographics Conference on Visualization — STARs, Borgo R., Maciejewski R., Viola I., (Eds.), EuroVis '14, The Eurographics Association*, pp. 125-139. <https://bit.ly/2Y3YeDu>
- Wanner, F., Weiler, A., & Schreck, T. (2012). Topic Tracker: Shape based visualization for trend and sentiment tracking in Twitter. *Proceedings of the 2nd IEEE Workshop on Interactive Visual Text Analytics “Task-Driven Analysis of Social Media”, TextVis '12*. <https://bit.ly/2AS6nCa>
- Wecker, A., Lanir, J., Mokryn, O., Minkov, E., & Kuflik, T. (2014). Semantize: Visualizing the sentiment of individual document. *Proceedings of the 2014 International Working*

- Conference on Advanced Visual Interfaces, AVI '14, ACM, pp. 385-386.
<https://bit.ly/2V91ons>
- Weiler, A., Grossniklaus, M., & Ssholl, M. (2015). The Stor-e-Motion visualization for topic evolution tracking in text data streams. *Proceedings of the 6th International Conference on Information Visualization Theory and Applications, IVAPP '15*, pp. 29-39.
<https://doi.org/10.5220/0005292900290039>
- Wensel, A., & Sood, S. O. (2008). VIBES: Visualizing changing emotional states in personal stories. *Proceedings of the 2nd ACM International Workshop on Story Representation, Mechanism and Context, SRMC '08, ACM*, pp. 49-56.
<https://doi.org/10.1145/1462014.1462025>
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05, Association for Computational Linguistics*, pp. 347- 354. <https://doi.org/10.3115/1220575.1220619>
- Wonderflow (2018). *10 Sentiment Analysis Examples That Will Help Improve Your Products*. [Online]. Ανάκτηση 21 Ιουλίου 2020, από: <https://bit.ly/3g1LfZz>
- Wu, J., Fu, Z., Liu, Z., Pan, J., Long, H., Lin, X., He, H., Chen, X., & Tang, J. (2012). City Flow: Prototype exploration for visualizing urban traffic conversations. *Proceedings of the International Conference on Privacy, Security, Risk and Trust and the International Conference on Social Computing, Social-Com/PASSAT 2012*, pp. 481-489.
<https://bit.ly/2BiId40>
- Wu, J., Fu, Z., Liu, Z., Lin, X., Tang, J., Pan, J., & Zhao, C. (2013). Creating reflections in public emotion visualization: Prototype exploration on traffic theme. *Proceedings of the 9th ACM Conference on Creativity & Cognition, C&C '13, ACM*, pp. 357-361.
<https://doi.org/10.1145/2466627.2466671>
- Wu, Y., Liu, S., Yan, K., Liu, M., & Wu, F. (2014). OpinionFlow: Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1763-1772. <https://doi.org/10.1109/TVCG.2014.2346920>
- Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., & Qu, H. (2010). OpinionSeer: Interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1109-1118. <https://doi.org/10.1109/TVCG.2010.183>
- Xu, K., Liao, S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for competitive intelligence. *Decision Support Systems*, 50(4), 743-754.
<https://doi.org/10.1016/j.dss.2010.08.021>
- Yatani, K., Novati, M., Trusty, A., & Truong, K. (2011). Review Spotlight: A user interface for summarizing user-generated reviews using adjective-noun word pairs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, ACM*, pp. 1541-1550. <https://bit.ly/2BiCQBS>

- Yu, Z., Wang, Z., Chen, L., Guo, B., & Li, W. (2016). Featuring, detecting, and visualizing human sentiment in Chinese micro-blog. *ACM Transactions on Knowledge Discovery from Data*, 10(4), 48:1-48:23. <https://doi.org/10.1145/2821513>
- Zhang, C., Liu, Y., & Wang, C. (2013). Time-space varying visual analysis of micro-blog sentiment. *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction, VINCI '13, ACM*, pp. 64-71. <https://doi.org/10.1145/2493102.2493110>
- Zhang, J., Kawai, Y., Kumamoto, T., & Tanaka, K. (2009). A novel visualization method for distinction of web news sentiment. In *Proceedings of the 10th International Conference on Web Information Systems Engineering (WISE '09)*, Vossen G., Long D. D. E., Yu J. X., (Eds.), Vol. 5802 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 181-194. https://doi.org/10.1007/978-3-642-04409-0_22
- Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y.-R., Collins, C. (2014a). #FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1773-1782. <https://bit.ly/3eOja7a>
- Zhao, J., Dong, L., Wu, J., & Xu, K. (2012). MoodLens: An emoticon-based sentiment analysis system for Chinese tweets. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, ACM*, pp. 1528-1531. <https://doi.org/10.1145/2339530.2339772>
- Zhao, J., Gou, L., Wang, F., & Zhou, M. (2014b). PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media. *Proceedings of the IEEE Conference on Visual Analytics Science and Technology, VAST '14, IEEE*, pp. 203-212. <https://bit.ly/3e5Oyxx>
- Zhao, S., Tong, Y., Liu, X., & Tan, S. (2016a). Correlating Twitter with the stock market through non-Gaussian SVAR. *Proceedings of the Eighth International Conference on Advanced Computational Intelligence, ICACI 2016*, pp. 257-264. <https://doi.org/10.1109/ICACI.2016.7449835>
- Zhao, Y., Qin, B., Liu, T., Tang, D. (2016b). Social sentiment sensor: a visualization system for topic detection and topic sentiment analysis on microblog. *Multimedia Tools and Applications*, 75(15), 8843-8860. <https://doi.org/10.1007/s11042-014-2184-y>