

ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΟΣ ΑΠΟ ΚΕΙΜΕΝΑ, ΓΙΑ ΟΝΤΟΤΗΤΕΣ
ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΑΥΤΩΝ.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ ΦΟΙΤΗΤΗ

ΑΝΕΣΤΑΚΗ ΑΝΑΣΤΑΣΙΟΥ
Α.Μ 13301

ΕΠΙΒΛΕΠΩΝ: ΒΑΡΛΑΜΗΣ Η , ΕΠ. ΚΑΘΗΓΗΤΗΣ

ΑΘΗΝΑ, ΙΟΥΛΙΟΣ 2015

Περιεχόμενα

1	Εισαγωγή.....	4
1.1	Ανάλυση συναισθήματος	4
1.2	Το πρόβλημα της Ανάλυσης συναισθήματος:.....	4
1.3	Ερευνητικά Ερωτήματα.....	5
2	Σχετικές Μελέτες:	6
2.1	Ανασκόπηση τεχνικών ανάλυσης συναισθήματος	6
2.2	Γενικές Προκλήσεις	9
2.2.1	Στόχος της Ανάλυσης συναισθήματος	9
2.2.2	Η Εξέλιξη της εξόρυξης γνώμης	10
2.2.3	Από λέξεις- κλειδιά στις έννοιες	10
2.3	Μελλοντικές επεκτάσεις της έρευνας.....	11
2.4	Ανάλυση Συναισθήματος με πολλούς τρόπους.....	11
3	Ανάλυση συναισθήματος και εξαγωγή γνώμης για τα Ελληνικά.....	13
3.1	Data scrapping και DEiXTo	13
3.2	NiosTo	15
3.3	Palo Pro & OpinionBuster.....	16
3.4	Ανάλυση συναισθήματος από Ο.Π.Α. (AUEB)	17
4	Υλοποίηση.....	18
4.1	Δημιουργία του Dataset με χρήση του DEiXTO.....	18
4.2	Αξιολόγηση του NiosTo	19
4.2.1	Η μέθοδος του NiosTo.....	19
4.3	Αξιολόγηση των υπολοίπων εφαρμογών	30
5	Αξιολόγηση σχολίων από την πλευρά των επιχειρήσεων	37
5.1	Online Customer Reviews.....	37
5.2	Online advertising	37
5.3	Πως οι καταναλωτές κρίνουν τη ποιότητα (quality) ενός ehop	37
5.4	Τι επηρεάζει τα σχόλια των online πελατών	40
5.5	Ποιος γράφει σχόλια και γιατί;.....	40
5.6	Πως χρησιμοποιούν οι επιχειρήσεις τα σχόλια των πελατών:	41
6	Μελέτες περίπτωσης	43
6.1	Case study Public	43

6.1.1	Η προτεραιότητά του Public.....	43
6.1.2	Η ποικιλία των σχολίων των πελατών.....	43
6.1.3	Οι βασικοί πυλώνες.....	43
6.1.4	Διαδικασία Προσέλκυσης & Επιλογής Προσωπικού εξασφαλίζει τη βέλτιστη Εξυπηρέτηση Πελατών	45
6.1.5	Διαχείριση των Αιτημάτων των πελατών της εταιρίας η οποία υποστηρίζει την Εξυπηρέτηση Πελατών	45
6.1.6	Παραδείγματα Αιτημάτων – Σχολίων πελατών.....	46
6.2	Case study Κωτσόβολος.....	47
6.2.1	Ηλεκτρονικό κατάστημα	47
6.2.2	Η ποικιλία των σχολίων των πελατών.....	48
6.2.3	Customer Plan	49
7	Βιβλιογραφία.....	50

Περιεχόμενα Πινάκων

Εικόνα 1: Το κεντρικό workspace της εφαρμογής Palo Pro

Εικόνα 2: Εφαρμογή DEiXTo

Εικόνα 3: Συγκέντρωση και εξαγωγή αποτελεσμάτων μέσα από το DEiXTo

Εικόνα 4: Κεντρικό περιβάλλον προγράμματος NiosTo

Εικόνα 5: Ανάλυση σχολίων με τη χρήση του NiosTo

Εικόνα 6: Ανάλυση σχολίων και εξαγωγή αρνητικών λέξεων

Εικόνα 7: Ανάλυση σχολίων και εξαγωγή θετικών λέξεων

Εικόνα 8: Ταξινόμηση σχολίων με βάση τη πολικότητα

Εικόνα 9: Ταξινόμηση σχολίων με βάση τα δεδομένα του χρήστη

Εικόνα 10: Ανάλυση των δεδομένων με βάση τη πολικότητα

Εικόνα 11: Λίστα θετικών και αρνητικών λέξεων. Ταξινόμηση με βάση τη πολικότητα.

Εικόνα 12: Ανάλυση δεδομένων από τα manual δεδομένα

Εικόνα 13: Ταξινόμηση σχολίων με βάση τα δεδομένα του χρήστη

Εικόνα 14: Γραφική παράσταση αλγόριθμου ταξινόμησης

Εικόνα 15: Πίνακας σύγκρισης αποτελεσμάτων προγραμμάτων aueb – Palo Pro

Εικόνα 16: Πραγματική ανάλυση με βάση τη πολικότητα

Εικόνα 17: Πίνακας σύγκρισης αποτελεσμάτων προγραμμάτων aueb – Palo Pro

Εικόνα 18: Συγκριτικός πίνακας δύο συστημάτων με τις πραγματικές τιμές

Εικόνα 19: Σχόλια πελατών από Reevoo για τον πελάτη public.gr

Εικόνα 20: Μηνιαία παρακολούθηση Customer Plan για την Κωτσόβολος

1 Εισαγωγή

Τα τελευταία χρόνια υπάρχει ραγδαία ανάπτυξη του Διαδικτύου. Όλο και περισσότεροι άνθρωποι ασχολούνται με την αναζήτηση περιεχομένου στο διαδίκτυο, την αγορά και πώληση προϊόντων καθώς επίσης και με την ανταλλαγή απόψεων για προϊόντα και υπηρεσίες που αγοράζουν. Οι χρήστες του Διαδικτύου δεν είναι πλέον παθητικοί αποδέκτες πληροφοριών. Μεγάλο ποσοστό χρηστών χρησιμοποιεί κοινωνικά δίκτυα και διάφορα forum για να ανταλλάξει απόψεις και να αξιολογήσει διάφορα προϊόντα και υπηρεσίες. Καθώς ολοένα και περισσότεροι χρήστες του Διαδικτύου αναρτούν κριτικές με τα προϊόντα ή τις υπηρεσίες που χρησιμοποιούν σε κοινωνικά δίκτυα, forum και ηλεκτρονικά καταστήματα πλέον υπάρχει πολύ μεγάλος όγκος πληροφοριών σε ηλεκτρονική μορφή σχετικά με τα συναισθήματα των ανθρώπων. Η ανάγκη ανάλυσης και κατά συνέπεια αξιοποίησης του όγκου της πληροφορίας αυτής οδήγησε στην εμφάνιση της *Ανάλυσης Συναισθήματος (Sentiment Analysis)*.

1.1 Ανάλυση συναισθήματος

Η *ανάλυση συναισθήματος* σε δεδομένα μεγάλου όγκου χρηστών του διαδικτύου κερδίζει ολοένα και μεγαλύτερο έδαφος στο ακαδημαϊκό χώρο καθώς και στον επιχειρηματικό χώρο. Οι ακαδημαϊκοί ερευνητές εντοπίζουν μεγάλο ενδιαφέρον στις τεχνικές προκλήσεις που παρουσιάζει η ανάλυση συναισθήματος ενώ, οι επιχειρηματίες εντοπίζουν το ενδιαφέρον τους στις πολλά υποσχόμενες προοπτικές της. Καινοτόμες επιχειρήσεις ασχολούνται με την εξόρυξη γνώσης μέσα από τις αξιολογήσεις χρηστών σε ηλεκτρονικά καταστήματα και κοινωνικά δίκτυα με βάση την ανάλυση συναισθήματος. Τέλος, υπάρχει μεγάλη επιρροή καταναλωτών και χρηστών του διαδικτύου από τις κριτικές που υπάρχουν στον ηλεκτρονικό κόσμο πριν λάβουν κάποια απόφαση για αγορά μιας υπηρεσίας ή ενός προϊόντος.

1.2 Το πρόβλημα της Ανάλυσης συναισθήματος:

Από τις πρώτες μελέτες που έγιναν από ερευνητές στον τομέα της Ανάλυσης συναισθήματος [1] είχε γίνει αντιληπτό ότι η κατηγοριοποίηση συναισθήματος (sentiment classification) διαφέρει από την απλή κατηγοριοποίηση κειμένου.

«Η κατηγοριοποίηση κειμένου – γνωστή και ως ταξινόμηση κειμένου ή ανίχνευση θέματος – αναφέρεται στην αντιστοίχιση κειμένων φυσικής γλώσσας σε θεματικές κατηγορίες ή κλάσεις οι οποίες ανήκουν σε ένα προκαθορισμένο σύνολο»[2]

Οι κατηγορίες, στην κατηγοριοποίηση συναισθήματος, καθορίζονται με βάση τα θέματα στόχους του εκάστοτε προβλήματος. Το πλήθος των κατηγοριών ποικίλει. Μπορεί να αποτελείται από ένα μικρό σύνολο 1,2 κατηγοριών αλλά μπορεί και να είναι ένα μεγάλο σύνολο. Για παράδειγμα οι κατηγορίες που υπάρχουν για την ταξινόμηση ενός άρθρου ενημερωτικού site με βάση τη θεματολογία του. Οι κατηγορίες που υπάρχουν είναι Αθλητικά, Οικονομικά, Επικαιρότητα κτλ. Ένα άρθρο μπορεί να ανήκει σε πάνω από μία κατηγορίες.

Αντίθετα, η Ανάλυση Συναισθήματος αναφέρεται σε ένα μικρό σύνολο κατηγοριών για παράδειγμα: θετικό, αρνητικό, ουδέτερο, 1 αστέρι... 5 αστέρια. Οι κατηγορίες αυτές είναι ανεξάρτητες της θεματολογίας του προβλήματος αλλά και αμοιβαία αποκλειόμενες. Το πρόβλημα που προσπαθεί να επιλύσει η Ανάλυση Συναισθήματος είναι ένα από τα πιο απλά προβλήματα με τα οποία ασχολείται η Επεξεργασία Φυσικής Γλώσσας [3]. Ένας υπολογιστής δεν είναι απαραίτητο να καταλαβαίνει πλήρως την σημασιολογία της κάθε λέξης αλλά θα πρέπει να εντοπίζει τη συνολική στάση του ανθρώπου που γράφει ένα σχόλιο ή μία κριτική. Οι παραπάνω απαιτήσεις απλοποιούν σε μεγάλο βαθμό τις απαιτήσεις κατανόησης της Φυσικής Γλώσσας από τους υπολογιστές αλλά δεν παύει να είναι μεγάλο πρόβλημα η ανίχνευση της πολικότητας, στο κείμενο, ακόμη και για τον άνθρωπο. [4]

1.3 Ερευνητικά Ερωτήματα

Σκοπός της έρευνάς μας θα είναι η σύγκριση και η ανάλυση συστημάτων που πραγματοποιούν εξόρυξη γνώμης και ανάλυση συναισθήματος μέσα από σχόλια που αφήνουν οι πελάτες σε ένα ηλεκτρονικό κατάστημα. Πιστεύουμε πως το να έχει κάποια εταιρία ένα πρόγραμμα που θα παρακολουθεί σε πραγματικό χρόνο τα σχόλια που αφήνουν οι πελάτες της και ταυτόχρονα την ανάλυσή τους σε αρνητικά, θετικά και ουδέτερα είναι χρήσιμο και αναγκαίο.

Αυτό που θα ερευνήσουμε είναι η σχέση μεταξύ της εξόρυξης γνώσης από σχόλια χρηστών και πως αυτά βοηθούν μια επιχείρηση λιανικού εμπορίου να αυξήσει την αφοσίωση των πελατών της (customer loyalty), τι μπορεί να βελτιώσει σε επίπεδο τιμών, πως μπορεί να αυξήσει την εμπειρία του πελάτη της (customer experience).

2 Σχετικές Μελέτες:

2.1 Ανασκόπηση τεχνικών ανάλυσης συναισθήματος

Παρά το γεγονός πως η ανάλυση συναισθήματος & η εξόρυξη γνώμης έχει παρουσιάσει ένα τεράστιο ενδιαφέρον για έρευνα, υπήρξε ένα υπόγειο ρεύμα ενδιαφέροντος για αρκετό διάστημα. Θα μπορούσε κανείς να μετρήσει πρώιμα έργα σχετικά με το αντικείμενο ως πρόδρομους της εποχής [10,11]. Αργότερα η έρευνα επικεντρώθηκε κυρίως στην ερμηνεία της μεταφοράς, της αφήγησης, της άποψης και συναφείς τομείς [12,13,14,15,16,17]. Το 2001 σηματοδοτείται η έναρξη της ευρείας ευαισθητοποίησης των ερευνητών γύρω από την ανάλυση συναισθήματος [18,19,20,21,22,23].

Οι παράγοντες που βοήθησαν στην εξέλιξη αυτή είναι:

- Η άνοδος των μεθόδων μηχανικής μάθησης (machine learning) σε φυσική γλώσσα
- Η διαθεσιμότητα των δεδομένων, για μεθόδους μηχανικής μάθησης, να εκπαιδευτούν και να αναπτυχθούν στην κριτική συσσώματωσης website
- Η υλοποίηση εφαρμογών νοημοσύνης

Η ανάλυση συναισθήματος & η εξόρυξη γνώμης έχουν σημαντικό ρόλο ως τεχνολογίες γενικής εφαρμογής και για άλλα συστήματα. Μια χρησιμότητα είναι σε συστήματα σύστασης (recommendation systems) [24,25] όπου όταν λαμβάνει ένα σύστημα πολλά αρνητικά σχόλια δεν είναι εφικτό να προταθεί. Επίσης χρησιμοποιείται στην ανίχνευση των απειλών στο ηλεκτρονικό ταχυδρομείο όταν κατηγοριοποιείται η επικοινωνία με το θέμα. Τέλος, ένα άλλο σημείο που χρησιμοποιείται η ανάλυση συναισθήματος είναι σε ηλεκτρονικές σελίδες όπου εμφανίζονται διαφημίσεις σε πλαϊνές μπάρες είναι χρήσιμο για την ανίχνευση ιστοσελίδων που περιέχουν ευαίσθητο περιεχόμενο ακατάλληλο για τοποθέτηση.

Οι **Ganu, Elhadad, Marian** [2009] [5] πραγματοποιούν μια μελέτη σχετικά με τη καλύτερη αξιολόγηση των προϊόντων χρησιμοποιώντας κείμενα καταναλωτών αντίθετα από αξιολογήσεις με αστέρια. Στόχος τους είναι να αναγνωρίσουν την πληροφορία που υπάρχει στα ελεύθερα και μη δομημένα σχόλια πελατών με σκοπό να βελτιώσουν την ακρίβεια των αποτελεσμάτων. Συγκεκριμένα η έρευνα εστιάζεται στη βελτίωση της προτροπής κάποιου να επισκεφθεί κάποιο εστιατόριο με μεγαλύτερη ακρίβεια. Οι ερευνητές ασχολούνται με την κατηγοριοποίηση και την ανάλυση συναισθήματος στο επίπεδο της πρότασης καθώς τα σχόλια που υπάρχουν στο διαδίκτυο μεταφέρουν πολλές φορές λεπτομερείς πληροφορίες σε λίγες μόνο λέξεις. Το project των ερευνητών το ονομάζουν URSA (User Review Structure Analysis). Με αυτό θέλουν να δημιουργήσουν καλύτερα εργαλεία για την αναγνώριση και ανάλυση των σχολίων των χρηστών.

Αρχικά τα δεδομένα είναι από το Citysearch New York και αφορούν 50000 κριτικές σε εστιατόρια. Πραγματοποιείται μία ανάλυση στα δεδομένα για να γίνει η

αναγνώριση των κατηγοριών και των συναισθημάτων καθώς και ο σχολιασμός ενός ταξινομημένου data set. Προτείνουν νέες μετρικές πρόβλεψης όπου λαμβάνουν υπόψη τους τις κριτικές που είναι γραπτές φράσεις. Γι' αυτό μετέφρασαν το κατηγοριοποιημένο data set σε κείμενο το οποίο έχει αξιολογηθεί. Αυτό το κείμενο χρησιμοποιείται για να γίνονται προβλέψεις. Τα αποτελέσματα της έρευνας αυτής δείχνουν ότι υπάρχουν καλύτερες κριτικές και προβλέψεις των εστιατορίων όταν οι σχολιασμοί των χρηστών είναι κείμενο παρά αν είναι αριθμητικές κριτικές (π.χ 1 Αστήρι.... 5 Αστέρια).

Οι **Παυλόπουλος και Ανδρουτσόπουλος** [6] στην έρευνά τους το 2014 στηρίχθηκαν στην ανάλυση των reviews των χρηστών με βάση το προϊόν προς αξιολόγηση. Η τεχνική με την οποία ασχολούνται ονομάζεται ABSA (Aspect-Based Sentiment Analysis). Η τεχνική ABSA χρησιμοποιείται για κάθε ξεχωριστό αντικείμενο (πχ. Μπαταρία) μιας οντότητας (πχ. Κινητό τηλέφωνο). Διαφορετικές λέξεις ή φράσεις μπορεί να χρησιμοποιηθούν για να αναφερθούν στο ίδιο αντικείμενο. Οι ερευνητές χωρίζουν την έρευνα σε δύο φάσεις. Στην πρώτη φάση όπου η σημασιολογική ομοιότητα των αντικειμένων είναι ορατή και η δεύτερη φάση που τα αντικείμενα είναι χωρισμένα είναι η καλύτερη επιλογή.

Σε ένα ABSA σύστημα δίνονται ένα πλήθος αξιολογήσεων για ένα προϊόν ή υπηρεσία (reviews από υπολογιστή). Το σύστημα προσπαθεί να αναγνωρίσει το κυριότερο χαρακτηριστικό του αντικειμένου (πχ motherboard) και το μέσο όρο σχολιασμού (πχ. 1 έως 5 αστέρια). Στην έρευνα αυτή οι ερευνητές χρησιμοποιούν 3710 σχόλια χρηστών στην Αγγλική γλώσσα από το Ganu et al (2009). Υπάρχουν επίσης σχόλια από φορητούς υπολογιστές 3085 σχολιασμοί από 394 πελάτες. Σαν αποτέλεσμα των ερευνών καταλήγουν στο ότι ένας καλός διαχωρισμός των χαρακτηριστικών ενός προϊόντος ή μιας υπηρεσίας με τη μέθοδο ABSA μπορεί να παρουσιάσει σημαντικά αποτελέσματα.

Οι **Minqing Hu και Bing Liu** [7] (2004) πραγματοποιούν την έρευνά τους πάνω στην εξόρυξη γνώσης και συγκέντρωση από διάφορες κριτικές πελατών. Καθώς το ηλεκτρονικό εμπόριο ολοένα και μεγαλώνει οι κριτικές, θετικές & αρνητικές, που λαμβάνει κάποιο προϊόν αυξάνονται με μεγάλο ρυθμό. Για παράδειγμα σε ένα δημοφιλές προϊόν οι κριτικές των πελατών είναι από εκατοντάδες μέχρι και χιλιάδες. Έτσι καταλαβαίνουμε πως υπάρχει μεγάλη δυσκολία σε κάποιον που θέλει να αγοράσει κάποιο προϊόν να διαβάσει όλες τις κριτικές και να αποφασίσει. Ακριβώς το ίδιο πρόβλημα αλλά σε μεγαλύτερο βαθμό υπάρχει και για τις ίδιες τις εταιρίες που εμπορεύονται κάποιο προϊόν το οποίο πωλείται σε πάνω από ένα site.

Σε αυτή την έρευνα οι ερευνητές έχουν σαν κύριο σκοπό την εξόρυξη γνώσης και τη συγκέντρωση των κριτικών διάφορων πελατών για ένα προϊόν. Πιο

συγκεκριμένα συγκεντρώνουν τα χαρακτηριστικά των προϊόντων που αξιολογήθηκαν από τους αγοραστές και στη συνέχεια αναγνωρίζουν με βάση το συναίσθημα σε κάθε μία κριτική αν είναι θετική ή αρνητική. Τέλος συγκεντρώνουν όλα τα αποτελέσματα μαζί.

Οι Ποντίκη, Γαλάνης, Παυλόπουλος, Παπαγεωργίου, Ανδρουτσόπουλος και Manandhar [8] πραγματοποίησαν το 2014 μια έρευνα πάνω στο ABSA (Aspect Based Sentiment Analysis) όπου την χωρίζουν σε τέσσερα επίπεδα. Στα δύο πρώτα επίπεδα τα δεδομένα της έρευνας για δύο κατηγορίες προϊόντων (εστιατόρια και φορητοί υπολογιστές) υπάρχουν ενώ στο τρίτο και τέταρτο επίπεδο της έρευνας υπάρχουν τα δεδομένα μόνο για τα εστιατόρια. Το πρώτο επίπεδο της έρευνας αφορά το Aspect term extraction. Μέσα από ένα σύνολο δεδομένων από αξιολογήσεις χρηστών για τα εστιατόρια και τους φορητούς υπολογιστές βρίσκουν όλες τις λέξεις κλειδί (aspects) που χαρακτηρίζουν το προϊόν ή την υπηρεσία. Αναγνωρίζουν σε αυτό το επίπεδο όλα τα aspects ακόμη και αυτά που έχουν ουδέτερη σημασία (ούτε θετικά ούτε αρνητικά). Στο δεύτερο επίπεδο της έρευνας ασχολούνται με το aspect term polarity. Χρησιμοποιούν τα aspects του πρώτου επιπέδου και τα κατηγοριοποιούν με βάση την πόλωσή τους (θετικά, αρνητικά, ουδέτερα). Στο τρίτο επίπεδο της έρευνάς τους ασχολούνται με το aspect category detection. Υπάρχει σε αυτό το επίπεδο ένα προκαθορισμένο σύνολο aspect, όπως είναι η τιμή και το φαγητό, και ένα σύνολο από σχόλια χρηστών. Να σημειώσουμε πως τα σχόλια χρηστών είναι χωρίς να έχουν κατηγοριοποιηθεί με βάση τη πολικότητά τους. Τέλος στο τέταρτο μέρος της έρευνας οι ερευνητές μελετούν πάνω στο aspect category polarity. Εδώ οι κατηγορίες για κάθε σχόλιο χρήστη παρέχονται. Σκοπός του επιπέδου αυτού είναι να καθορίσουν τη πολικότητα των reviews (θετική, αρνητική, ουδέτερη) για κάθε κατηγορία aspect που συζητείται σε κάθε σχόλιο. Τα δεδομένα που χρησιμοποίησαν για τα σχόλια των χρηστών στην κατηγορία εστιατόρια ήταν από το Ganu (2009) και αφορούν 3041 σχόλια ενώ στη κατηγορία των φορητών υπολογιστών χρησιμοποίησαν 3845 σχόλια.

Οι Suresh, Roodi και Ειρηνάκη [9] το 2014 στην έρευνά τους ασχολούνται με ένα σύστημα το οποίο εξάγει aspects με ακριβής χαρακτηρισμούς από σχόλια χρηστών και προτείνει σχόλια σε άλλους χρήστες με βάση τα πρότυπα αξιολόγησης. Οι ερευνητές σε αυτή την έρευνα προτείνουν ένα σύστημα που παράγει αυτόματα εξατομικευμένες προτάσεις σχολίων στους χρήστες χρησιμοποιώντας δύο διαφορετικές προσεγγίσεις. Αρχικά υπάρχουν τα παραδοσιακά συστήματα φιλτραρίσματος όπου το σύστημα παράγει τη βαθμολόγηση των προφίλ του κάθε χρήστη και προσαρμόζει την λίστα με τις αξιολογήσεις του καθενός χρήστη. Η δεύτερη προσέγγιση που απασχολεί τους ερευνητές έχει να κάνει με την εξόρυξη γνώμης ώστε να προσδιορίσει τα σημαντικά χαρακτηριστικά σε κάθε κριτική. Το σύστημα που αναλύουν στην συγκεκριμένη έρευνα έχει σύνολο δεδομένων από το Yelp.

Συγγραφέας	Επιστημονικό Υπόβαθρο	Συνεισφορά	Αποτελέσματα
Ganu, Elhadad, Marian	User Review Structure Analysis	Δεδομένα από Citysearch NY	Καλύτερα εργαλεία ανάλυσης σχολίων
Παυλόπουλος, Ανδρουτσόπουλος	Aspect Based Sentiment Analysis	Δεδομένα από Ganu et al (2009)	Σημαντικά αποτελέσματα με κατάλληλο διαχωρισμό των χαρακτηριστικών
Minqing Hu, Bing Liu	Ανάλυση συναισθήματος	Ηλεκτρονικά καταστήματα	Συγκέντρωση και σύγκριση αποτελεσμάτων
Ποντίκη, Γαλάνης, Παυλόπουλος, Παπαγεωργίου, Ανδρουτσόπουλος και Manandhar	Aspect term extraction	Δεδομένα από Ganu (2009)	Καθορισμός πολικότητας των σχολίων
Suresh, Roodi και Ειρηνάκη	Ανάλυση συναισθήματος	Δεδομένα από Yelp	Εξόρυξη γνώμης από σχόλια

Συγκεντρωτικός πίνακας μελετών

2.2 Γενικές Προκλήσεις

Το αναπτυσσόμενο ερευνητικό ενδιαφέρον στην ανάλυση συναισθήματος & στην εξόρυξη γνώμης οφείλεται στις εφαρμογές που αναφέρθηκαν προηγουμένως. Ποια είναι η διαφορά της ανάλυσης συναισθήματος από το κλασσικό text mining όμως. Η κατηγοριοποίηση κατά κείμενο (text mining) ταξινομεί τα έγγραφα από το θέμα. Σε αντίθεση με την κατηγοριοποίηση συναισθήματος (sentiment classification) υπάρχουν διαφορετικές κλάσεις (θετικό ή 3 αστέρια) που γενικεύονται σε πολλές περιοχές και χρήστες.

Υπάρχουν πολλά χαρακτηριστικά απαντήσεων σε ερωτήσεις που σχετίζονται με τη γνώμη όπου διαφέρουν από τις ερωτήσεις που έχουν ως βάση το γεγονός [26]. Αυτό έχει σαν αποτέλεσμα την εξαγωγή της opinion – oriented πληροφορίας να διαφέρει από την παραδοσιακή εξαγωγή πληροφορίας [27].

2.2.1 Στόχος της Ανάλυσης συναισθήματος

Ο βασικός στόχος της εξόρυξης γνώμης είναι η κατηγοριοποίηση της πολικότητας (polarity classification). Η κατηγοριοποίηση της πολικότητας συμβαίνει όταν ένα κομμάτι κειμένου διατυπώνει γνώμη σχετικά με ένα θέμα. Κριτικές χρηστών όπως «like», «dislike» είναι παραδείγματα της πολικότητας ταξινόμησης. Η ταξινόμηση της πολικότητας προσδιορίζει τις υπέρ και κατά εκφράσεις σε σχόλια χρηστών στο διαδίκτυο κάνοντας τις αξιολογήσεις του προϊόντος περισσότερο

αξιόπιστες. Η ανίχνευση συμφωνίας (agreement detection) είναι μια άλλη μορφή κατηγοριοποίησης συναισθήματος. Καθορίζει αν ένα ζεύγος από έγγραφα κειμένου με ίδιο ή διαφορετικό συναισθημα θα λάβουν ίδια ετικέτα.

2.2.2 Η Εξέλιξη της εξόρυξης γνώμης

Αυτή τη στιγμή η εξόρυξη γνώμης και η ανάλυση συναισθήματος βασίζονται στην εξαγωγή διανυσμάτων (vectors) που εκπροσωπούν τα πιο σημαντικά χαρακτηριστικά στο κείμενο. Μπορούμε να χρησιμοποιήσουμε αυτά τα διανύσματα για να ταξινομήσουμε τα πιο σχετικά χαρακτηριστικά. Θα μπορούσαμε να πούμε πως δύο συχνά χρησιμοποιούμενες λειτουργίες είναι η συχνότητα (frequency) και η παρουσία (presence). Η παρουσία είναι ένα δυαδικό χαρακτηριστικό στο οποίο οι εγγραφές γράφουν μόνο αν η τιμή είναι 1 ή δε γράφουν αν η τιμή είναι 0. Η παρουσία (presence) αποτελεί μια πιο αποτελεσματική βάση για την αναθεώρηση της κατηγοριοποίησης της πολικότητας.

2.2.3 Από λέξεις- κλειδιά στις έννοιες

Μπορούμε να μελετήσουμε την εξέλιξη της ανάλυσης συναισθήματος από τα αναλυτικά διακριτικά (tokens) ή δομικά στοιχεία και τις πληροφορίες που σχετίζονται με αυτά τα διακριτά. Μπορούμε να κατηγοριοποιήσουμε τις υπάρχουσες προσεγγίσεις σε τέσσερις κατηγορίες. Η επισήμανση των λέξεων κλειδιών, η λεξιλογική συγγένεια, οι στατιστικές μέθοδοι και οι τεχνικές που βασίζονται στην έννοια.

Η επισήμανση των λέξεων: αν και μπορεί να χαρακτηριστεί ως η πιο αφελής προσέγγιση η οικονομία και η προσβασιμότητα του εντοπισμού των λέξεων κλειδιών την καθιστούν δημοφιλή. Αυτή η προσέγγιση κατατάσσει κείμενο ανάλογα με τις λέξεις κλειδιά: χαρούμενος, λυπημένος, φοβισμένος, βαρετά κτλ.

Η λεξιλογική συγγένεια: Αυτή η προσέγγιση δεν ανιχνεύει προφανείς λέξεις αλλά αποδίδει λέξεις με πιθανή συγγένεια με συγκεκριμένα συναισθήματα. Για παράδειγμα η λέξη «ατύχημα» παρουσιάζει κατά 75% μια αρνητική επιρροή. Αυτή η προσέγγιση ξεπερνά την προσέγγιση επισήμανση των λέξεων αλλά έχει και τα μειονεκτήματά της. Αρχικά αναιρεί εκφράσεις της μορφής «Απέφυγα ένα ατύχημα» καθώς θα την εμφανίσει κατά 75% αρνητική και δεν καταλαβαίνει γλωσσικές ιδιομορφίες της γλώσσας (I met my girlfriend by accident).

Οι στατιστικές μέθοδοι: Είναι σημασιολογικά αδύναμες που σημαίνει ότι έχουν μικρή προγνωστική αξία. Αυτό έχει ως αποτέλεσμα η προσέγγιση αυτή να λειτουργεί σωστά μόνο όταν λαμβάνει επαρκώς μεγάλη εισαγωγή κειμένου. Έτσι θα μπορούσαμε να πούμε πως οι στατιστικές μέθοδοι ενώ είναι σε θέση να ταξινομούν με βάση το συναισθημα το κείμενο ενός χρήστη σε επίπεδο σελίδας δεν λειτουργεί καλά σε μικρότερα κείμενα όπως προτάσεις και σχόλια χρηστών.

Οι τεχνικές βασισμένες στην έννοια: Οι τεχνικές αυτές χρησιμοποιούν οντολογίες διαδικτύου ή σημασιολογικά δίκτυα ώστε να επιτευχθεί σημασιολογική ανάλυση.

2.3 Μελλοντικές επεκτάσεις της έρευνας

Αυτή τη στιγμή τα κοινωνικά δίκτυα, τα forum, τα διάφορα site προσφέρουν στους χρήστες νέες μεθόδους για να μοιράζει ο κάθε χρήστης περιεχόμενο και πληροφορίες με τους φίλους του. Η πληροφορία αυτή επειδή προέρχεται από τον άνθρωπο είναι αδόμητη και δεν μπορεί η μηχανή – ο υπολογιστής να την επεξεργαστεί. Η δυνατότητα που μπορεί να έχει ο άνθρωπος να συλλέγει πληροφορίες που υπάρχουν στο διαδίκτυο για τη γνώμη των ανθρώπων ως προς τις κοινωνικές εκδηλώσεις, τη πολιτική αστάθεια, στρατηγικές μάρκετινγκ των εταιριών και για τα προϊόντα που εμπορεύονται οι επιχειρήσεις παρουσιάζει μεγάλο ενδιαφέρον στην επιστημονική κοινότητα και στο κόσμο των επιχειρήσεων. Το αποτέλεσμα αυτής της ανάγκης είναι η εξόρυξη γνώμης και η ανάλυση συναισθήματος (opinion mining και sentiment analysis). Τόσο η εξόρυξη γνώμης όσο και η ανάλυση συναισθήματος επικεντρώνονται στη αναγνώριση της πολικότητας και στην αναγνώριση του συναισθήματος. Τα δύο αυτά ερευνητικά πεδία συχνά χρησιμοποιούνται ως συνώνυμα καθώς χρησιμοποιούν εξόρυξη γνώσης (data mining) και τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) για να ανακαλύψουν και να παρουσιάσουν πληροφορίες και γνώμες μέσα από τις αμέτρητες πληροφορίες που δίνει το διαδίκτυο.

Η εξόρυξη γνώμης μέσα από φυσικές γλώσσες είναι δύσκολη καθώς απαιτεί μια βαθιά κατανόηση του συντακτικού και των σημασιολογικών κανόνων που έχει η γλώσσα. Για πολλές εταιρίες η ανάλυση συναισθήματος και η εξόρυξη γνώμης αποτελούν μέρος της έρευνάς τους. Για παράδειγμα πολλές εταιρίες χρησιμοποιούν την εξόρυξη γνώμης στα εταιρικά ηλεκτρονικά τους καταστήματα με σκοπό να έχουν άμεση πληροφόρηση για τη γνώμη και την ικανοποίηση των πελατών τους. Τα συστήματά τους συγκεντρώνουν συνέχεια μεγάλο όγκο πληροφοριών από το διαδίκτυο όπως αναλύσεις προϊόντων, σχόλια για το brand τους και για θέματα πολιτική. Άλλα συστήματα των εταιριών χρησιμοποιούν την εξόρυξη γνώμης και την ανάλυση συναισθήματος για να βελτιώσουν την πελατειακή τους σχέση και να λάβουν θετικά και αρνητικά σχόλια για τα προϊόντα και τις υπηρεσίες τους.

Οι επιχειρήσεις χρησιμοποιούν την ανάλυση συναισθήματος για να δημιουργήσουν στρατηγικές μάρκετινγκ έχοντας πρόσβαση και προβλέποντας τις συμπεριφορές του κοινού τους αναφορικά με το brand τους. Η έρευνα και η ανάπτυξη των εταιριών ασχολείται με τη δημιουργία crawlers όπου συλλέγουν σχόλια χρηστών από το διαδίκτυο.

2.4 Ανάλυση Συναισθήματος με πολλούς τρόπους

Ολοένα και περισσότερες νέες πηγές για εξόρυξη γνώμης και ανάλυσης συναισθήματος δημιουργούνται. Οι κάμερες που υπάρχουν στα «έξυπνα» κινητά τηλέφωνα, στα tablet και στις άλλες συσκευές επιτρέπουν στους χρήστες να δημοσιεύουν τις γνώμες τους σε μορφή ήχου ή βίντεο περισσότερο απ' ότι σε κείμενο. Σκεφτείτε την ποσότητα του υλικού που δημοσιεύει ο άνθρωπος μέσα σε ένα λεπτό στο youtube και η οποία είναι 48 ώρες βίντεο κάθε ένα λεπτό της ώρας.

Πολλοί νέοι τομείς έρευνας μπορεί να είναι χρήσιμοι στην εξόρυξη γνώμης, όπως η έκφραση του προσώπου, η κίνηση του σώματος ή ένα βίντεο ενός χρήστη.

Ο παγκόσμιος ιστός (Web) αρχικά λειτουργούσε ως «μόνο για ανάγνωση» (read only) και αυτή τη στιγμή έχει εξελιχθεί σε «ανάγνωση – γραφή» (read- write). Αυτή η εξέλιξη δημιούργησε ενθουσιώδης χρήστες που αλληλεπιδρούν μεταξύ τους και διαδίδουν γνώση μέσα από τα κοινωνικά δίκτυα, τις διαδικτυακές κοινότητες και τα blogs. Συλλογική γνώση έχει εξαπλωθεί σε ολόκληρο το διαδίκτυο και ιδιαίτερα σε τομείς που σχετίζονται με την καθημερινή ζωή όπως το εμπόριο, τουρισμός και εκπαίδευση. Παρά τη σημαντική αυτή πρόοδο η εξόρυξη γνώσης και η ανάλυση συναισθήματος εξακολουθούν να είναι στο επίκεντρο των επιστημονικών ερευνών. Πρόσφατες προσεγγίσεις στοχεύουν στην καλύτερη κατανόηση των εννοιολογικών κανόνων που διέπουν το συναίσθημα.

Τα μελλοντικά συστήματα εξόρυξης γνώμης χρειάζονται ευρύτερη και βαθύτερη γνώση βάσης. Περισσότερη ολοκληρωμένη γνώση που πρέπει να συνδυάζεται με την ανθρώπινη σκέψη και ψυχολογία. Αυτό θα οδηγήσει σε μία καλύτερη κατανόηση της φυσικής γλώσσας και θα γεφυρώσει το χάσμα μεταξύ των αδόμητων πληροφοριών που δίνει ο άνθρωπος και των δομημένων πληροφοριών που χρειάζεται ένας υπολογιστής για να τις επεξεργαστεί. Συνδυάζοντας επιστημονικές θεωρίες των συναισθημάτων με την πρακτική της μηχανικής θα οδηγηθούμε σε διαφορετικές προσεγγίσεις για τον σχεδιασμό ευφών συστημάτων εξόρυξης γνώμης.

3 Ανάλυση συναισθήματος και εξαγωγή γνώμης για τα Ελληνικά

Καθημερινά πολλές εταιρίες οι οποίες ασχολούνται με τις πωλήσεις προϊόντων ή υπηρεσιών και επενδύουν μεγάλα ποσά στην έρευνα και την ανάπτυξη. Ένα κομμάτι το οποίο τα τελευταία χρόνια αναπτύσσεται με μεγάλους ρυθμούς είναι η συγκέντρωση πληροφοριών μέσα από διάφορες ιστοσελίδες και blog. Εκεί ο καταναλωτής μπαίνει και γράφει ελεύθερα τη γνώμη του και τους προβληματισμούς του. Υπάρχουν διάφορα προγράμματα τα οποία κάνουν data scrapping για την εξαγωγή των κειμένων στην Αγγλική γλώσσα. Ο όγκος της πληροφορίας που εξάγουν τέτοια λογισμικά είναι τεράστιος και χρειάζεται η απαραίτητη αποκωδικοποίηση της πληροφορίας. Για την Ελληνική γλώσσα το λογισμικό που χρησιμοποίησα στην έρευνα μου ήταν το DEiXTo.

Όπως αναφέραμε ο όγκος της πληροφορίας που συλλέγουν διάφορα προγράμματα data scrapping είναι μεγάλος και θα πρέπει να γίνει με προσοχή η ανάλυση των αποτελεσμάτων ώστε να συλλέξουμε την πληροφορία που έχει αξία για κάθε επιχείρηση. Η ανάλυση συναισθήματος (sentiment analysis) έρχεται στην συνέχεια να μας προβληματίσει αναφορικά με την Ελληνική γλώσσα. Η Ελληνική γλώσσα είναι γεμάτη με λέξεις οι οποίες μπορεί να δηλώνουν μία ή και παραπάνω έννοιες. Επίσης άλλη δυσκολία που παρουσιάζεται σε αυτά τα συστήματα είναι όταν γίνεται η ανάλυση μιας φράσης η οποία περιέχει επίθετο (πχ. Ωραίος) καθώς υπάρχει περίπτωση με την τοποθέτηση μιας λέξης μπροστά από το θετικό επίθετο αυτό να μετατραπεί σε αρνητικό. Στη έρευνα μου χρησιμοποίησα προγράμματα τα οποία κάνουν ανάλυση συναισθήματος σε κείμενα με την Ελληνική γλώσσα όπως το Palo Pro καθώς και ένα ερευνητικό πρόγραμμα του Ο.Π.Α.

3.1 Data scrapping και DEiXTo

Το DEiXTo είναι ένα πρόγραμμα το οποίο μας επιτρέπει να εξάγουμε περιεχόμενο από ιστοσελίδες. Το πρόγραμμα αυτό δουλεύει σε ένα γραφικό περιβάλλον μέσω του οποίου ο χρήστης ορίζει κανόνες εξαγωγής (extraction rules) και το πρόγραμμα εξάγει τα δεδομένα από τις ιστοσελίδες.

Αρχικά ο χρήστης μπορεί στη πάνω αριστερή πλευρά του προγράμματος να γράψει το URL της html ιστοσελίδας που επιθυμεί να χρησιμοποιήσει. Το πρόγραμμα ελέγχει αυτόματα την ιστοσελίδα και σε λίγα δευτερόλεπτα έχει κάνει την ανάλυση του κώδικα html. Το DEiXTo δεν υποστηρίζει σελίδες που έχουν πλαίσια (frames).

Η λειτουργία του DEiXTo βασίζεται στο **μοντέλο αντικειμένου εγγράφου (Document object model – DOM)**. Έτσι λοιπόν πάνω δεξιά στο πρόγραμμα υπάρχει η οπτικοποίηση της DOM αναπαράστασης της ιστοσελίδας. Η ανάλυση αυτή γίνεται με χρήση των διεπαφών που προσφέρει το DOM και του αλγορίθμου σάρωσης δέντρου κατά βάθος (depth first). Για κάθε κόμβο του DOM αποθηκεύονται σημαντικά στοιχεία της ιστοσελίδας όπως υπερσύνδεσμοι, εικόνες, κείμενο και

πηγαίο κώδικα HTML. Για να γίνει πιο απλοποιημένη η DOM αναπαράσταση της ιστοσελίδας οι δημιουργοί του DEiXTo χρησιμοποιούν την λειτουργία «ignore html tags» όπου στην ανάλυση που χρησιμοποιούμε μπορούμε να αποκλείσουμε κάποιους τύπους HTML κόμβους (tag filtering) όπως είναι bold, italic, strong κτλ.

Μόλις ολοκληρωθεί η λήψη της ιστοσελίδας και η ανάλυση DOM ο χρήστης πρέπει να ορίσει κάποιο κανόνα εξαγωγής. Ακουμπώντας το κέρσορα σε οποιοδήποτε σημείο της ιστοσελίδας που έχει κάνει ανάλυση ο χρήστης εμφανίζεται κίτρινο πλαίσιο σε κάθε κατηγορία της σελίδας. Το επόμενο βήμα είναι δεξί κλικ και επιλέγουμε use a record instance. Υλοποιούμε το κανόνα αυτό πατώντας execute rule και στη συνέχεια save as. Τότε μας κάνει εξαγωγή το πρόγραμμα ένα txt – XML - RSS αρχείο με τα δεδομένα που έχουμε επιλέξει εμείς. Εύκολα αυτό το αρχείο το μεταφέρουμε σε excel αρχείο για να το κάνουμε εισαγωγή σε οποιοδήποτε πρόγραμμα επιθυμούμε. Τα αρχεία εξόδου περιλαμβάνουν κωδικοποίηση UTF-8 ανεξάρτητα από την κωδικοποίηση της κάθε ιστοσελίδας.

Σε κάποιες περιπτώσεις που η δομή του προτύπου είναι σχετικά απλή τότε έχουμε σαν συνέπεια να επιστρέφονται λανθασμένα υπεράριθμα αποτελέσματα. Το πρότυπο τότε πρέπει να γίνει πιο αυστηρό και απαιτείται να περιγραφεί το περιβάλλον ή αλλιώς η γειτονιά του ριζικού κόμβου του κανόνα. Στο DEiXTo αυτό το επιτυγχάνεται με εισαγωγή κανόνα ορισμένων άμεσων προγόνων του τρέχοντος ριζικού κόμβου του κανόνα. Ο χρήστης μπορεί να ανεβοκατεβαίνει επίπεδα και να προσθέτει αδέρφια (siblings) σε ένα κόμβο πρόγονο της αρχικής ρίζας.

Ο αλγόριθμος μόλις διαβάσει την ιστοσελίδα εντοπίζει σε ορισμένες περιπτώσεις διαδοχικούς κόμβους που αποτελούν μία ομάδα. Αυτό σημαίνει ότι πηγαίνουν πάντα μαζί οπότε και ο αλγόριθμος τους χειρίζεται ως ομάδα κόμβων. Στο DEiXTo υπάρχει η επιλογή FSON (Following Successive Optional Nodes) όπου εκφράζει το πλήθος των επόμενων διαδοχικών προαιρετικών κόμβων.

Ο αλγόριθμος συμφωνίας προτύπου (pattern matching) που χρησιμοποιεί το DEiXTo είναι αποτελεσματικός για τη πλειοψηφία των περιπτώσεων. Για τον εντοπισμό στιγμιότυπων επιθυμητής πληροφορίας της ιστοσελίδας γίνεται προσπέλαση κάθε στοιχείου της DOM αναπαράστασης. Σε κάθε DOM κόμβο αντιστοιχεί ένας νέος κύκλος προσπάθειας σύμφωνα με το πρότυπο. Ο αλγόριθμος αποτελείται από δύο βασικά στάδια. Αρχικά ελέγχεται αν το υποδένδρο ταιριάζει με το υποδένδρο του κανόνα με ρίζα το κόμβο vroot (vroot, κόμβος της εικονικής ρίζας του δένδρου του κανόνα) και στο δεύτερο στάδιο ελέγχεται αν η γειτονιά του κόμβου DOM υπό εξέταση ταιριάζει με τη γειτονιά του vroot.

Η κεντρική ιδέα πίσω από τον αλγόριθμο που χρησιμοποιεί το DEiXTo είναι πως για να ταιριάζουν δύο κόμβοι πρέπει να έχουν την ίδια ετικέτα και να ταιριάζουν και οι θυγατρικοί τους κόμβοι. Οπότε το ταίριασμα δύο δέντρων ανάγεται σε πρόβλημα ταιριασματος των ριζικών τους κόμβων όπου επιτυγχάνεται μέσω αναδρομής (recursion) και σάρωσης του δέντρου κατά βάθος. Τα βασικά

χαρακτηριστικά του αλγόριθμου είναι η υποστήριξη ενδεχόμενης απουσίας κόμβων (missing nodes) στο δέντρο στόχο και η ύπαρξη προαιρετικών τμημάτων στον κανόνα.

Υπάρχει η δυνατότητα ο χρήστης να αποθηκεύσει τον κανόνα εξαγωγής που έχει χρησιμοποιήσει με σκοπό να τον χρησιμοποιήσει και αργότερα. Έτσι δε θα είναι αναγκασμένος να φτιάχνει τους ίδιους κανόνες εξαγωγής ξανά και ξανά για τις ίδιες σελίδες. Για το σκοπό αυτό υλοποιήθηκε λειτουργία αποθήκευσης ενός wrapper, όπου παρέχεται η δυνατότητα εκτέλεσής του ανά πάσα στιγμή με εντολή του χρήστη. Όλες οι απαραίτητες πληροφορίες για τον wrapper αποθηκεύονται σε ένα XML αρχείο ώστε να μπορεί ο χρήστης να φορτώσει τον wrapper. Τα XML αρχεία αυτά ονομάζονται wrapper project files και έχουν wpf κατάληξη.

Το DEiXTo έχει δυνατότητα να παράγει έξοδο σε RSS αρχείο. Τα item στοιχεία του channel δημιουργούνται από τα δεδομένα που εξάγονται από τα στιγμιότυπα επιθυμητής πληροφορίας που εντοπίστηκαν.

3.2 NiosTo

Το NiosTo είναι ένα πρόγραμμα το οποίο το έχει σχεδιάσει η ομάδα που έχει σχεδιάσει και το DEiXTo. Είναι μια εφαρμογή που υλοποιεί ένα αλγόριθμο εξόρυξης γνώμης (opinion mining tool) καθώς επίσης κάνει και ταξινόμηση λέξεων με βάση το συναίσθημα. Υπάρχει άμεση σχέση του NiosTo με το DEiXTo καθώς το δεύτερο αναλύει μια ιστοσελίδα και εξάγει τα επιθυμητά δεδομένα σε αρχείο excel, ενώ το πρώτο δέχεται σαν είσοδο τα δεδομένα αυτά και προχωράει σε αναλύσεις.

Τα πλεονεκτήματα του αλγορίθμου που υλοποιεί το NiosTo είναι τα παρακάτω:

- Εισάγει μια νέα προσέγγιση για να δημιουργηθούν συγκεκριμένα λεξικά γνώμης (opinion dictionaries)
- Παρέχει μια πειραματική μελέτη όπου όλα τα βήματα του προτεινόμενου αλγορίθμου αξιολογούνται μέσα από ένα εργαλείο ανάλυσης συναισθήματος.
- Παρέχει ένα σύνολο δεδομένων με απόψεις στην ελληνική γλώσσα.

Το NiosTo λαμβάνει σαν είσοδο αρχεία csv και παρουσιάζει τις λέξεις που ανακαλύπτει σε κάθε επίπεδο. Επιπλέον παρουσιάζει με γραφική απεικόνιση τα αποτελέσματα καθώς επίσης και την κατηγοριοποίηση συναισθήματος. Όλοι οι παράμετροι μπορούν να αλλάζουν με το γραφικό περιβάλλον που παρουσιάζει το NiosTo.

Τέλος για να είναι εκτελέσιμο το NiosTo απαιτεί την ύπαρξη Java Runtime Environment (JRE).

3.3 Palo Pro & OpinionBuster

Η εφαρμογή Palo Pro είναι υπεύθυνη για τη γλωσσολογική ανάλυση κειμένων που περιέχουν συναίσθημα (αρνητικές ή θετικές αναφορές) για πρόσωπα εταιρίες και προϊόντα.

Είναι μία συνδρομητική υπηρεσία η οποία αθροίζει όλες τις ειδήσεις, blog, συζητήσεις και τα βίντεο στα ελληνικά μέσα και αναλύει τα δεδομένα με σκοπό τη διαχείριση της φήμης του πελάτη. Ο χρήστης έχει ένα φιλικό και χρήσιμο εργαλείο για τη παρακολούθηση της εικόνας της επιχείρησής του παρακολουθώντας σε πραγματικό χρόνο τα θετικά, αρνητικά ή ουδέτερα σχόλια που αναρτούνται. Εφόσον το Palo Pro είναι μια εμπορική υπηρεσία συνδρομής οι απαιτήσεις για ακριβή αποτελέσματα είναι υψηλές με σκοπό να έχει ακρίβεια πάνω από 85% και για τη πολικότητα των σχολίων αλλά και για την αναγνώριση των οντοτήτων. Τα δεδομένα που συλλέγονται από το Palo Pro καθαρίζονται και φιλτράρονται σε πραγματικό χρόνο και αυτό έχει σαν αποτέλεσμα να μπορεί ο πελάτης να δει άμεσα τα σχόλια που τον αφορούν. Κατά μέσο όρο σε μια ημέρα το Palo Pro έχει εντοπίσει πάνω από 2,5 εκατομμύρια σχόλια καθώς το πρόγραμμα σαρώνει 1400 διαφορετικά site. Αυτές οι ιστοσελίδες κατηγοριοποιούνται με βάση τη σημασία τους και με βάση το τομέα ειδήσεων. Τα σχόλια των χρηστών είναι συνήθως σε μη επίσημη γραφή, χωρίς συντακτικό και με διάφορα σύμβολα.

Το Palo Pro οργανώνεται γύρω από ένα χώρο εργασίας (workplace) μέσω του οποίου παρακολουθεί ο χρήστης σε πραγματικό χρόνο τη φήμη της εταιρίας του. Ο χρήστης μπορεί να δημιουργήσει ένα νέο χώρο εργασίας και μπορεί να επιλέξει ένα ή περισσότερα πρόσωπα, brands, κτλ.



Εικόνα 1: Το κεντρικό workspace της εφαρμογής Palo Pro

Το **OpinionBuster** είναι το προϊόν πάνω στο οποίο στηρίζεται το Palo Pro, καθώς είναι υπεύθυνο για την εξαγωγή των οντοτήτων και την πολικότητα που συνδέονται με αναφορές τους στα κείμενα. Για να μπορεί το **OpinionBuster** να έχει τα καλύτερα δυνατά αποτελέσματα ενσωματώνει στοιχεία της φυσικής γλώσσας που αποτελούνται από οντολογίες και αποκλείει συστήματα που βασίζονται σε αλγορίθμους μηχανικής μάθησης. Το **OpinionBuster** έχει αναπτυχθεί στο πλαίσιο της Ellogon3 (Πετάσης, 2002) σε μία γλώσσα μηχανικής που κωδικοποιήθηκε σε C/C++. Το Ellogon3 προσφέρει την απαιτούμενη υπολογιστική απόδοση για ένα εμπορικό προϊόν επιτυγχάνοντας ταχύτητα επεξεργασίας 100 σχόλια ανά δευτερόλεπτο, σε έναν επεξεργαστή Intel 3930K.

3.4 Ανάλυση συναισθήματος από Ο.Π.Α. (AUEB)

Το δεύτερο πρόγραμμα που μελετήθηκε για την ανάλυση των σχολίων και την κατηγοριοποίησή τους σε θετικά, αρνητικά και ουδέτερα είναι των Μακρυγιάννη και Βασσάλου [30] από το Οικονομικό Πανεπιστήμιο Αθηνών που το δημιούργησαν για τη πραγματοποίηση της έρευνάς τους : «Εξαγωγή συναισθήματος από Ελληνικά tweets».

Η προεργασία που γίνεται σε αυτό το πρόγραμμα είναι να αφαιρεθούν οι υπερσύνδεσμοι από τα σχόλια, οι αναφορές στα άτομα (@user), τα hashtags, η αφαίρεση των αρνητικών και θετικών emoticons με τα ☺ ☹. Η πολικότητα των σχολίων σε αυτό το σύστημα χαρακτηρίζεται ως, θετικό, αρνητικό, ουδέτερο, θετικό προς αρνητικό, αρνητικό προς θετικό. Σε σχετικό πείραμα που πραγματοποίησαν οι ερευνητές συγκέντρωσαν 3191 Ελληνικά tweets από τα οποία 973 θετικά, 1450 αρνητικά και 768 ουδέτερα. Το σύστημα χρησιμοποιεί ένα δείγμα κειμένων εκπαίδευσης για να εκπαιδεύσει έναν ταξινομητή με χρήση Support Vector Machines

.

4 Υλοποίηση

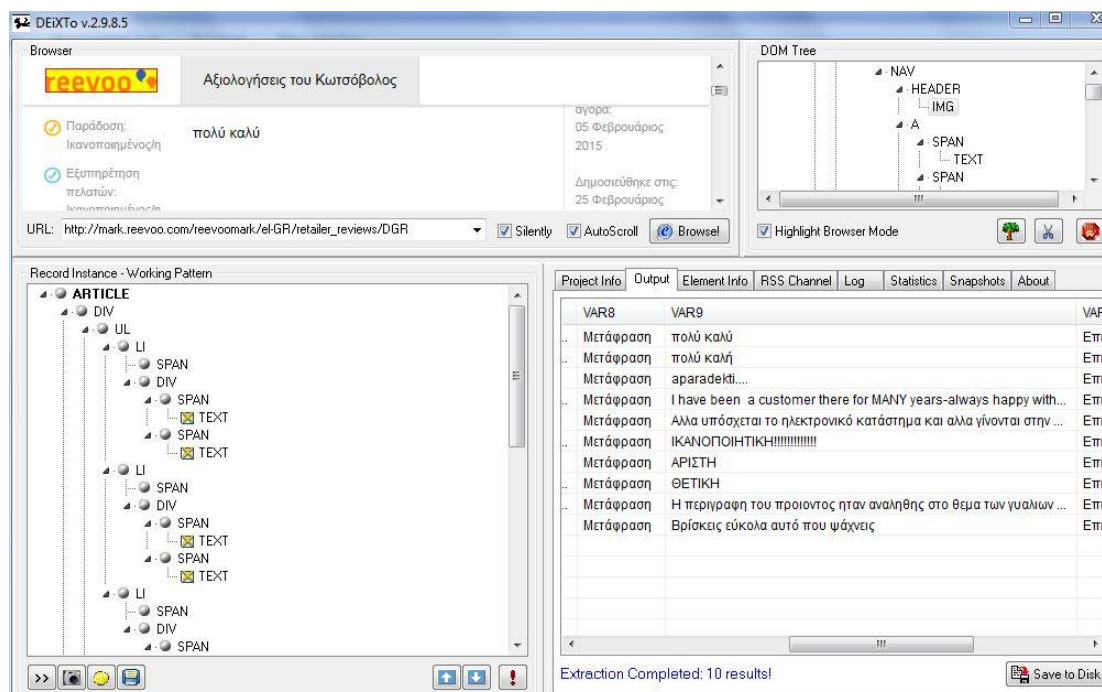
4.1 Δημιουργία του Dataset με χρήση του DEiXTO

Στη διαδικασία της ανάλυσης της χρησιμότητας του παραπάνω προγράμματος (DEiXTO) δοκίμασα να εξάγω τα σχόλια των πελατών που έχουν αγοράσει από τα ηλεκτρονικά καταστήματα δύο Retail εταιριών (www.kotsovolos.gr , www.public.gr)

Στο online κατάστημα της εταιρίας Κωτσόβολος υπάρχουν σχόλια πελατών σε εξωτερικά link (http://mark.reevoo.com/reevoo/el-GR/retailer_reviews/DGR) τα οποία είναι 470 σχόλια κειμένου.

Βήμα 1^ο:

Η ανάλυση της ιστοσελίδας στο DEiXTO και ο ορισμός κανόνα εξαγωγής.



Εικόνα 2: Εφαρμογή DEiXTO

Βήμα 2^ο:

Η αποθήκευση των αποτελεσμάτων ανά σύνδεσμο σε txt αρχείο

Αρχείο	Επεξεργασία	Μορφή	Προβολή	Βοήθεια
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Στέλιος Μετάφραση	μολοντι είμαι πελάτης χρόνια και έχω πραγματοποιήσει όλες μου τις αγο
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Θεοδώρας Μετάφραση	Γενικότερα αν και υπάρχει μια απειρία για τις μεγάλες αλυσίδες
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Ευάγγελος Μετάφραση	είμαι ευχαριστημένος από τις υπηρεσίες και τα προϊόντα, μέχρι
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Κωστής Μετάφραση	Δεν είχα ποτέ παραπονα για τον Κωτσόβολο.
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Σπύρος Μετάφραση	Η αγορά έγινε μέσω διαδικτύου. Η εξυπηρέτηση πελατών έλυσε άμεσα ένα ζ
υπόλο: Δυσχερστημένος/η	περιόδοι:	Ικανοποιημένος/η	Επιβεβαιωμένος αγοραστής	Μετάφραση Παναγιώτο
υπόλο: Δυσχερστημένος/η	περιόδοι:	Ικανοποιημένος/η	ΣΟΦΙΑ Μετάφραση	Όσον αφορά την παραγγελία μου από το διαδικτυο δεν αντιμετωπίσα κανένα
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Αλέξανδρος Μετάφραση	Εξαιρετική στις online αγορές στο τοπικό κατάστημα μετρία εξυπ
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	ΓΕΩΡΓΙΑ Μετάφραση	ΟΙ ΗΛΕΚΤΡΙΚΕΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ ΜΟΥ ΕΙΝΑΙ ΟΛΕΣ ΑΠΟ ΤΟΝ ΚΩΤΣΟΒΟΛΟ
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Αλεξάνδρος Μετάφραση	Αν και υπήρξε πρόβλημα στο ταμείο με την εξαργύρωση κουπονιού
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Επιβεβαιωμένος αγοραστής	Μετάφραση κανένα παραπονο...
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Επιβεβαιωμένος αγοραστής	Μετάφραση Συμφέρον, άμεσα εξυπηρέτηση
υπόλο: Ικανοποιημένος/η	περιόδοι:	Ικανοποιημένος/η	Θωμάς Μετάφραση	Αφγοί, αλλά έτσι για να είμαστε ειλικρινείς, βελτιώστε λίγο τ
υπόλο: Δυσχερστημένος/η	περιόδοι:	Δυσχερστημένος/η	Θωμάς Μετάφραση	ΑΠΑΡΑΔΕΚΤΗ ΕΞΥΠΗΡΕΤΗΣΗ
υπόλο: Ικανοποιημένος/η	περιόδοι:	Δυσχερστημένος/η	Μαρία Μετάφραση	Είμαι πελάτης στο κατάστημα πολλά χρόνια είμαι ευχαριστημένη εκτός απ

Εικόνα 3: Συγκέντρωση και εξαγωγή αποτελεσμάτων μέσα από το DEiXTO

Βήμα 3^ο:

Η συγκέντρωση των αποτελεσμάτων και για τα 30 διαφορετικά link που έχει στην ιστοσελίδα του ο Κωτσόβολος με τα σχόλια σε ένα αρχείο excel για μετέπειτα χρήση.

Η ίδια διαδικασία έγινε και για το δεύτερο online κατάστημα της εταιρίας Public (www.public.gr) όπου εκεί υπήρχαν 1400 σχόλια πελατών σε 155 διαφορετικά link.

4.2 Αξιολόγηση του Niosto

Το πρόγραμμα Niosto το έχουν δημιουργήσει οι Αγαθαγκέλου Π, Κατάκης Ι, Κόκκορας Φ και Ντόνας Κ. και έχει να κάνει με την επίλυση του προβλήματος της εξόρυξης γνώμης. Το Niosto μπορεί να δεχθεί ως είσοδο δεδομένα και γνώμες από τα social media, διάφορα blogs και forum. Στην συνέχεια πραγματοποιεί ανάλυση των δεδομένων με τη χρήση λεξικού λέξεων. Προκύπτει τέλος η κατηγοριοποίηση των κριτικών σε θετικές – αρνητικές και προβάλλει τα αποτελέσματα με γραφικές παραστάσεις.

Χαρακτηριστικά Λεξικού: Τα χαρακτηριστικά του λεξικού που χρησιμοποιεί το Niosto περιλαμβάνουν μια λίστα λέξεων των οποίων η πόλωση γνωρίζουμε εξαρχής ποια είναι (θετική ή αρνητική). Τέλος, για τις λέξεις γνωρίζουμε τη δύναμη των λέξεων και την ένταση του συναισθήματος που εκπέμπουν. Για παράδειγμα η λέξη «πανεμόρφος» έχει ένταση συναισθήματος 10 ενώ η λέξη «ψυχρός» έχει ένταση συναισθήματος -4.

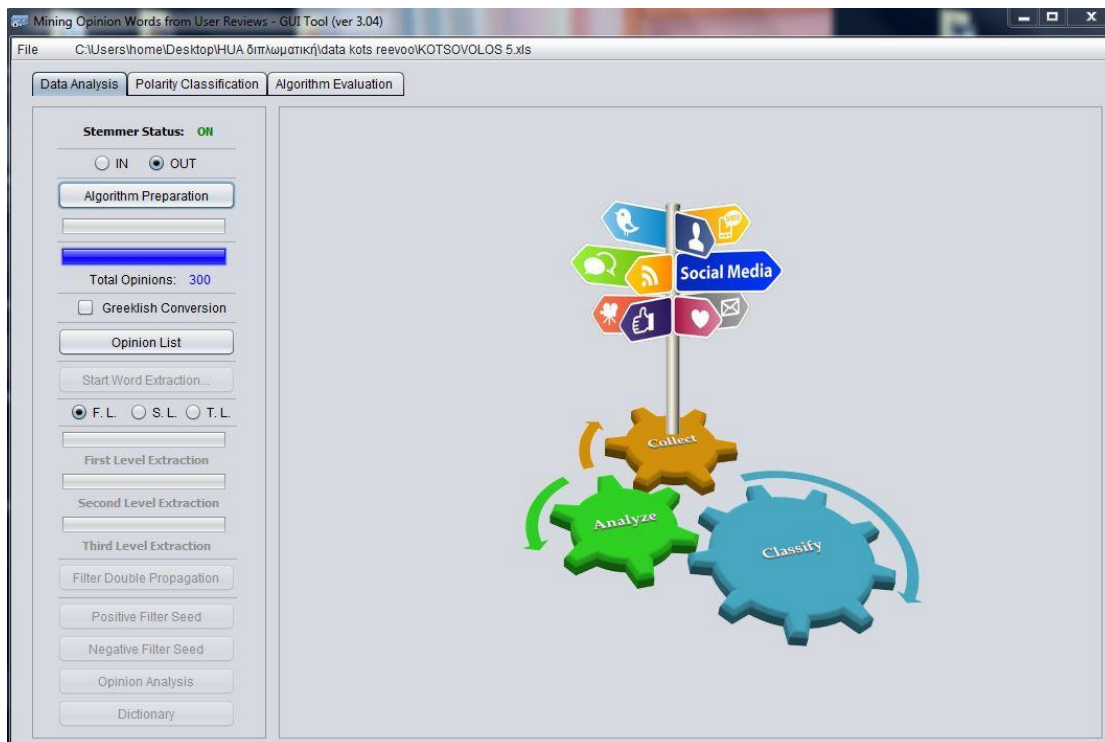
4.2.1 Η μέθοδος του Niosto

- Προεπεξεργασία γνώμης (opinion preprocessing): Το Niosto δέχεται ως είσοδο δεδομένα σε απλή μορφή όπου μπορεί να περιέχουν σύμβολα κτλ. (πχ Η μπαταρία # διαρκεί ελάχιστες ημέρες, σε εμένα!!) . Με την προεπεξεργασία αφαιρεί τα σύμβολα και λαμβάνει υπόψιν του το κόμμα και τη τελεία που δηλώνουν συναισθημα.
- Βοηθητική λίστα έτοιμων λέξεων (auxiliary list preparation): Υπάρχει λίστα με τα άρθρα, τα βασικά ρήματα, επίθετα σε συγκριτικό και υπερθετικό βαθμό, ουσιαστικά, αρνητικές λέξεις κτλ. Τέτοιες έτοιμες λίστες βοηθούν το Niosto στην ανάλυση των σχολίων.
- Εισαγωγή πληροφορίας και φιλτράρισμα πληροφοριών εξόδου. (Seed import and filtered seed extraction): Πραγματοποιεί ανάλυση σε επίπεδο προτάσεων με βάση τη βοηθητική λίστα έτοιμων λέξεων . (πχ. «Λίγο περισσότερο

χρήσιμος» άρα λίγο= μειώνει, περισσότερο= συγκριτικός βαθμός, χρήσιμος= θετικό) Διαθέτει το NiosTo 21 θετικούς τέτοιους συνδυασμούς και 12 αρνητικούς συνολικά.

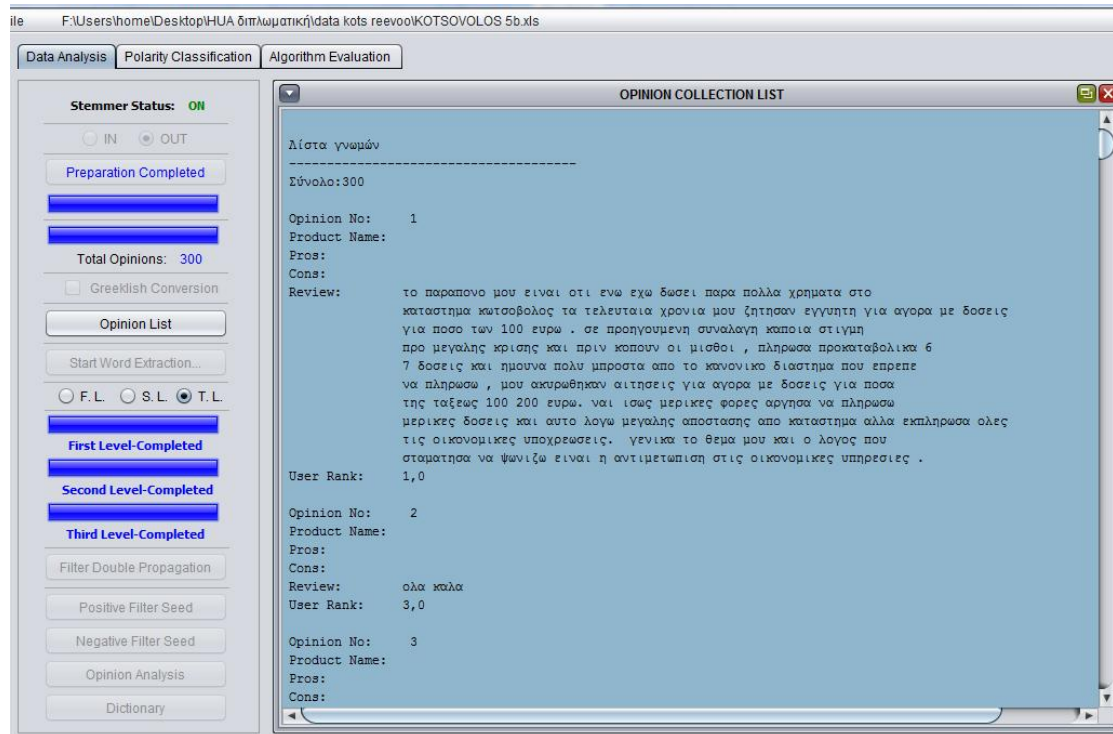
- Συνεργασία με βάση την εξαγωγή (Conjunction based extraction): Στο συγκεκριμένο στάδιο το NiosTo χρησιμοποιεί 6 θετικούς και 4 αρνητικούς συνδυασμούς εξαγωγής και βγάζει τα αποτελέσματά του.
- Διπλή διασπορά και επικύρωση λέξεων (Double propagation & opinion word validation): Στο πρώτο βήμα έχουμε την εσωτερική ανάλυση συναισθήματος για παράδειγμα στη φράση «Υπέροχη άνετη οθόνη» η λέξη «άνετη» βρίσκεται μπροστά από τη λέξη οθόνη και έχει πολικότητα συναισθήματος +1 . Στο δεύτερο βήμα έχουμε την εξωτερική ανάλυση συναισθήματος με όριο +3. Για παράδειγμα η φράση «εύκολη πλοήγηση όμορφα γραφικά» λαμβάνει πολικότητα +2.

Στο πρώτο παράδειγμα έχουμε εξάγει σχόλια χρηστών από την ιστοσελίδα της εταιρίας Κωτσόβολος με τη βοήθεια του DEiXTo και στη συνέχεια περάσαμε τα δεδομένα αυτά στο NiosTo. Παρακάτω υπάρχουν τα αποτελέσματα του προγράμματος για 300 σχόλια χρηστών.



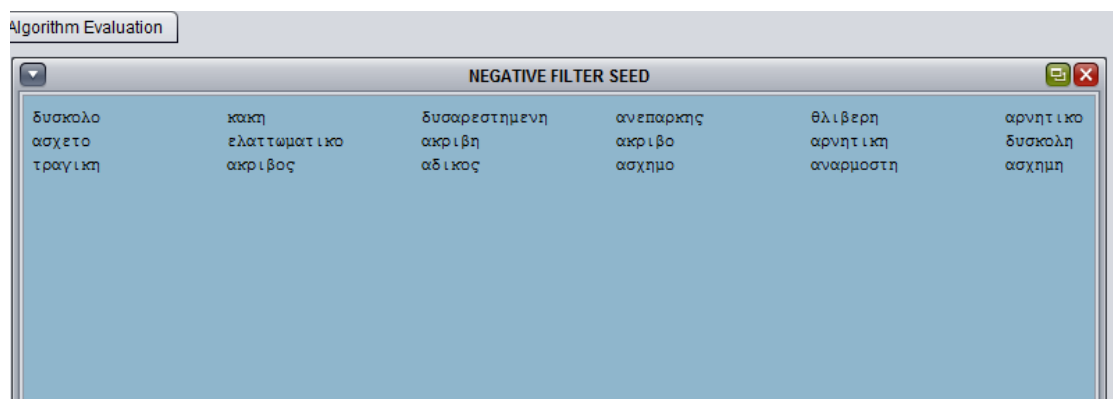
Εικόνα 4: Κεντρικό περιβάλλον προγράμματος NiosTo

Μόλις εισάγουμε τα δεδομένα του πρώτου ηλεκτρονικού καταστήματος που θα μελετήσουμε, η εφαρμογή NiosTo πραγματοποιεί την πρώτη ανάλυση των δεδομένων. Το πρόγραμμα διαβάζει ένα – ένα το σχόλια και πραγματοποιεί ανάλυση τριών επιπέδων.



Εικόνα 5: Ανάλυση σχολίων με τη χρήση του NiosTo

Σαν συνέχεια πραγματοποιεί διαχωρισμό των θετικών και αρνητικών λέξεων που εμφανίζονται μέσα στα σχόλια των χρηστών

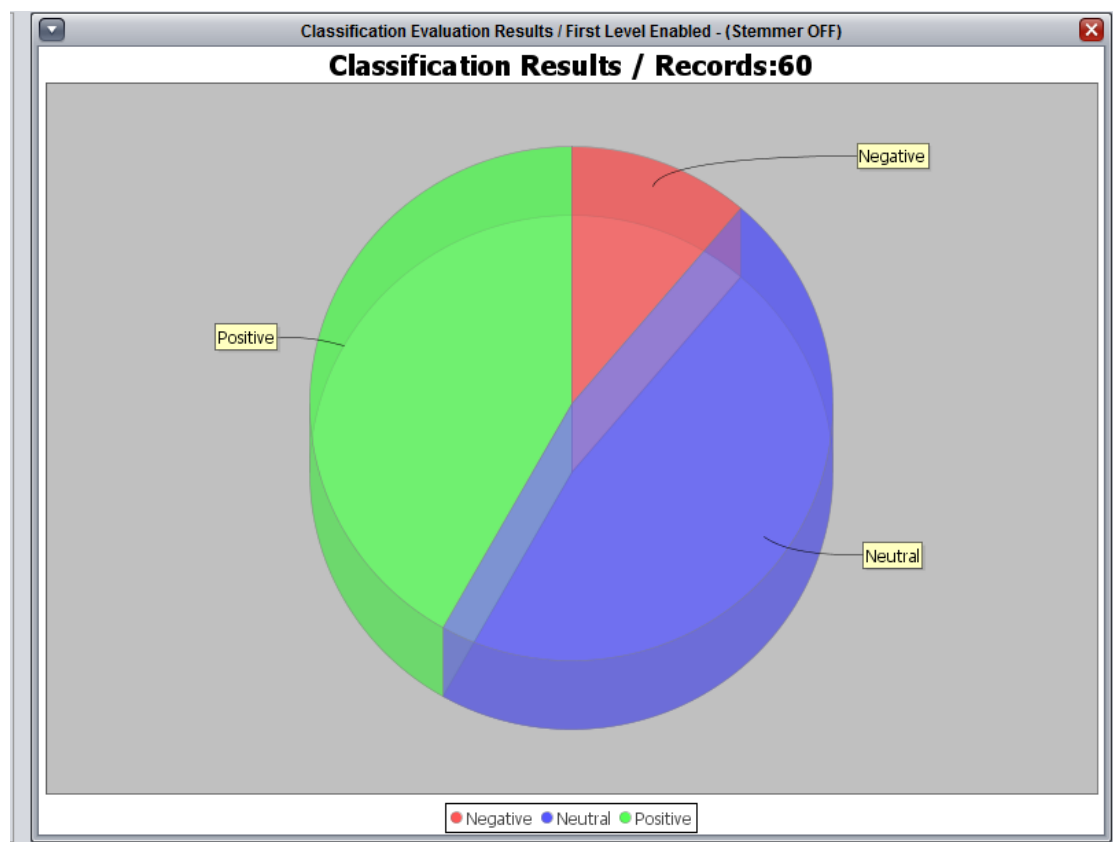


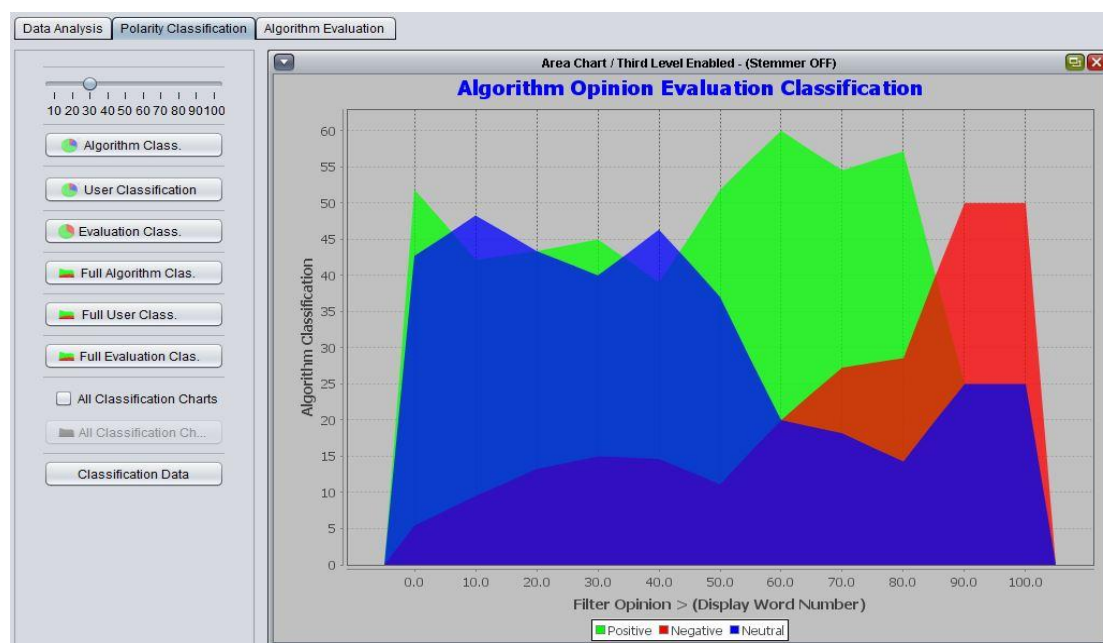
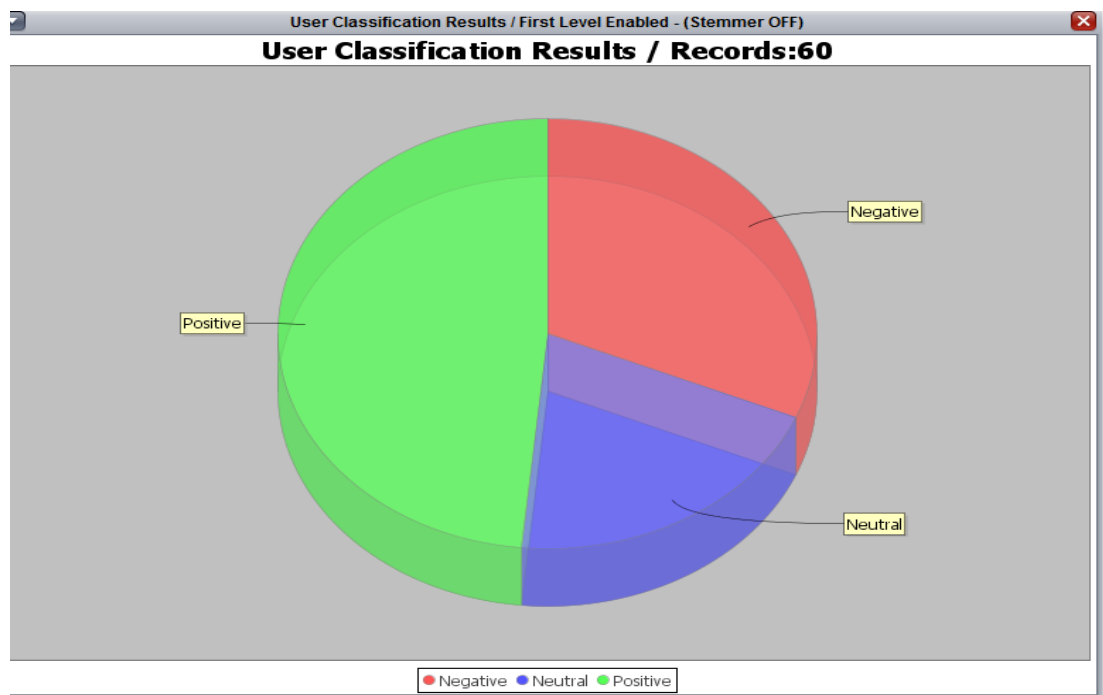
Εικόνα 6: Ανάλυση σχολίων και εξαγωγή αρνητικών λέξεων

POSITIVE FILTER SEED					
καλα	καλη	ωραια	αψογα	μεγαλη	βελτιωθει
αξιοπιστη	καλο	ακριβως	αριστη	εξαιρετικη	ακριβης
ολοκαινουργιο	ευχαριστημενη	αψογος	μεγαλο	ικανοποιητικη	υποδειγμα
γρηγορη	ευχαριστημενος	σταθερο	θετικο	σουπερ	βολικη
ευγενικη	προθυμη	δελεαστικη	θετικη	ευκολη	τελειος
τελεια	φιλικη	εξαιρετικος	υπευθυνο	ωραιο	σημαντικο
πιστος	αρμοδιο	υπεροχο	εξαιρετικο	κορυφαια	αξιοπρεπη
προστατευτικη	προσεκτικο	σημαντικη	καλος	υπεροχη	

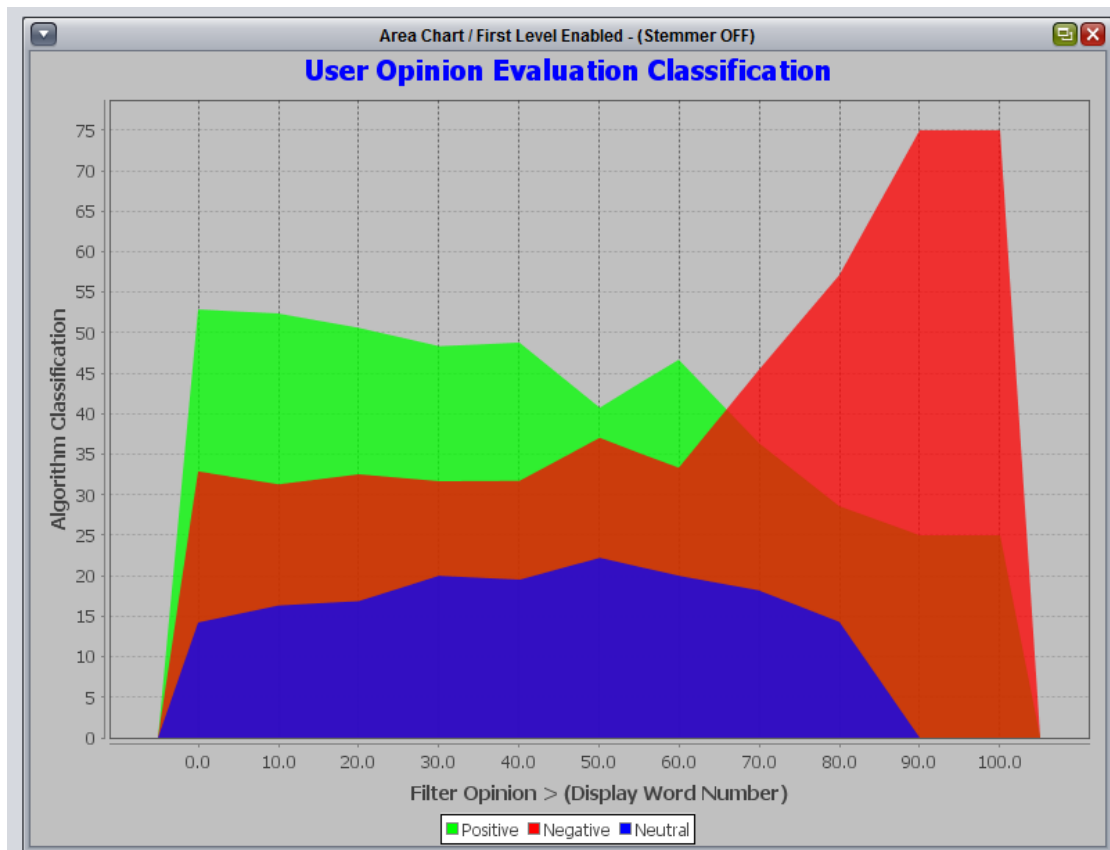
Εικόνα 7: Ανάλυση σχολίων και εξαγωγή θετικών λέξεων

Παρακάτω υπάρχει η σύγκριση στα screenshots από την ταξινόμηση με βάση την πολικότητα των σχολίων του NiosTo & της manual εκδοχής. Έτσι τα κατηγοριοποιεί το NiosTo σε θετικά, αρνητικά και ουδέτερα.





Εικόνα 8: Ταξινόμηση σχολίων με βάση τη πολικότητα



Εικόνα 9: Ταξινόμηση σχολίων με βάση τα δεδομένα του χρήστη

Με τη γραφική παράσταση στην εικόνα 9 συμπεραίνουμε πως για ένα μεγάλο εύρος τιμών της παραμέτρου Filter (word number 5-60%) οι τιμές της ακρίβειας κυμαίνονται ως εξής:

Accuracy

Positive: 0.40-0.52

Negative: 0.32-0.37

Neutral: 0.15-0.20

OPINION POLARITY CLASSIFICATION DATA									
Words >	Records	Alg Pos	Alg Neg	Alg Neutral	User Pos	User Neg	User Neutral	Match	No Match
0	295	51,19	4,41	44,41	52,88	32,88	14,24	33,56	66,44
10	147	40,82	7,48	51,70	52,38	31,29	16,33	31,97	68,03
20	83	40,96	9,64	49,40	50,60	32,53	16,87	28,92	71,08
30	60	41,67	11,67	46,67	48,33	31,67	20,00	30,00	70,00
40	41	36,59	9,76	53,66	48,78	31,71	19,51	29,27	70,73
50	27	48,15	7,41	44,44	40,74	37,04	22,22	29,63	70,37
60	15	60,00	13,33	26,67	46,67	33,33	20,00	33,33	66,67
70	11	54,55	18,18	27,27	36,36	45,45	18,18	27,27	72,73
80	7	57,14	14,29	28,57	28,57	57,14	14,29	14,29	85,71
90	4	25,00	25,00	50,00	25,00	75,00	0,00	0,00	100,00
100	4	25,00	25,00	50,00	25,00	75,00	0,00	0,00	100,00

Εικόνα 10: Ανάλυση των δεδομένων με βάση τη πολικότητα

Στην Εικόνα 10 παρατηρούμε πως η στήλη match το NiosTo για τα δεδομένα του παραδείγματος kotsonolos.gr μας εμφανίζει μέγιστο accuracy = 33,33 και γενικά οι τιμές του accuracy και για τις τρεις κατηγορίες είναι περίπου στο 30.

Στο δεύτερο παράδειγμα έχουμε εξάγει σχόλια χρηστών από την ιστοσελίδα της εταιρίας Public με τη βοήθεια του DEiXTo και στη συνέχεια περάσαμε τα δεδομένα αυτά στο NiosTo. Παρακάτω υπάρχουν τα αποτελέσματα του προγράμματος για 1332 σχόλια χρηστών.

File C:\Users\home\Desktop\HUA διπλωματική\data public reevoo\data public review mono 03.xls

Data Analysis Polarity Classification Algorithm Evaluation

Stemmer Status: **ON**

☐ IN ☒ OUT

Preparation Completed

Total Opinions: 1332

☐ Greeklish Conversion

Opinion List

Start Word Extraction...

☐ F.L. ☐ S.L. ☒ T.L.

First Level-Completed

Second Level-Completed

Third Level-Completed

Filter Double Propagation

Positive Filter Seed

Negative Filter Seed

Opinion Analysis

Dictionary

EXTRACTED OPINION DICTIONARY

Λεξικό εξορυγμένων λέξεων γνώμης (Opinion Mining Dictionary)

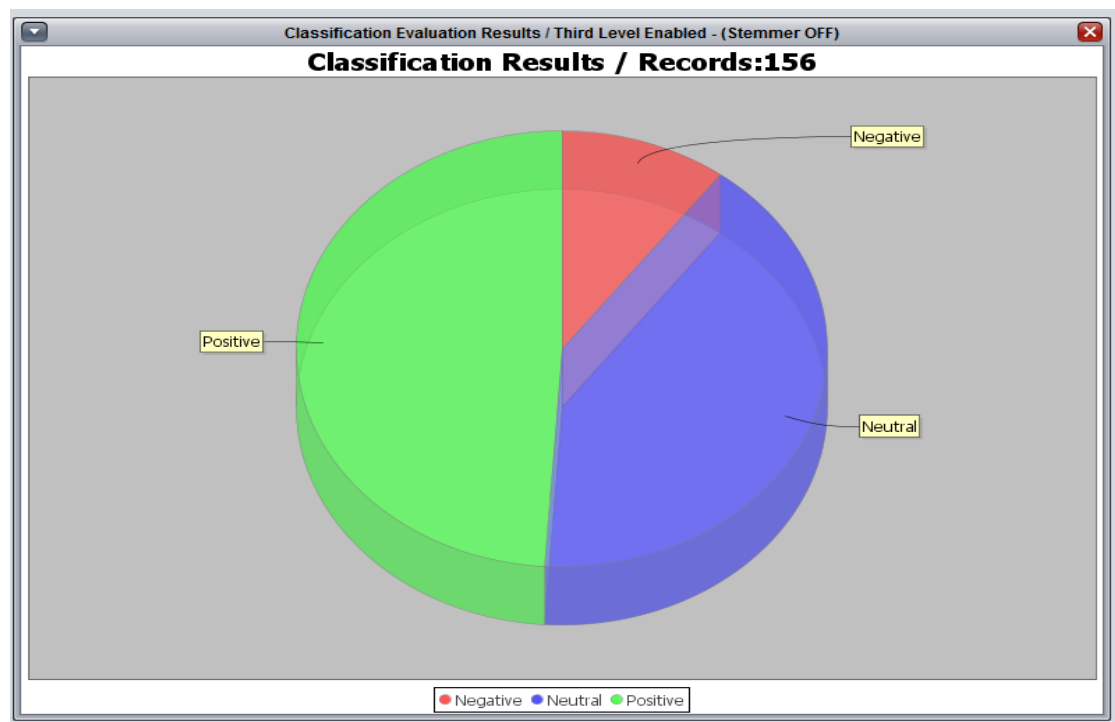
Λίστα θετικά Εξορυγμένων Λέξεων Γνώμης και Πολικότητα (Positive Filter Seed & Polarity)

Σύνολο: 97 λέξεις

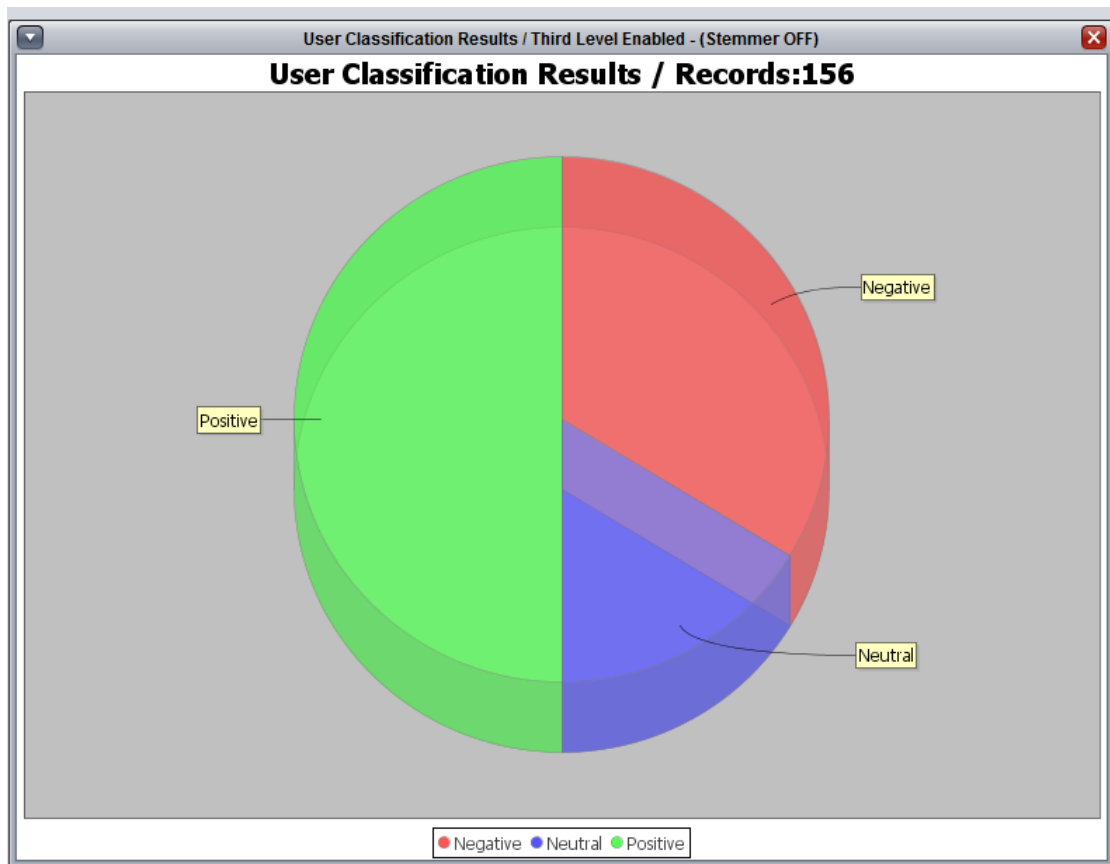
καλη	[243,0]	φιλικό	[12,0]	προθυμο	[2,0]	απολαυστική	[1,0]	μεγα
οικονομική	[3,0]	εξαιρετική	[51,0]	θετική	[12,0]	γρήγορη	[72,0]	αρισ
καλό	[47,0]	ευχαριστήμενος	[26,0]	ευγενική	[4,0]	τελεία	[33,0]	επιπλ
ικανοποιητική	[15,0]	αψογά	[13,0]	ευκολή	[22,0]	δυναμικό	[2,0]	αρισ
ευγενικό	[13,0]	ευχαρίστο	[3,0]	ιδανικός	[2,0]	αξιοπιστή	[8,0]	ευχαρι
ζεστο	[2,0]	καλά	[44,0]	μεγάλο	[5,0]	λογικό	[1,0]	τελεία
ευκολο	[12,0]	τεράστια	[5,0]	ευχαρίστη	[1,0]	αποδοτική	[1,0]	ακριβή
ωραίο	[1,0]	υπευθυνος	[1,0]	ιδανικό	[1,0]	ευχαριστος	[2,0]	ωραία
φοβερή	[4,0]	καλός	[3,0]	αποτελεσματικός	[1,0]	εξαιρετικό	[12,0]	χρησιμ
αυσιαστική	[3,0]	υπεροχή	[5,0]	πραγματική	[1,0]	καθαρή	[-1,0]	καθαρο
γρήγορο	[7,0]	ιδιαίτερη	[0,0]	μεγάλος	[2,0]	φωτεινό	[1,0]	απιστευ
ιδιαίτερο	[2,0]	ικανοποιητικό	[2,0]	φιλοξένο	[1,0]	απιστευτή	[1,0]	ολοκλή
ιδιαίτερος	[1,0]	σβαρη	[1,0]	προσεκτική	[1,0]	υπευθυνο	[1,0]	αξιοπισ
βολικό	[1,0]	λειτουργικό	[1,0]	φοβερο	[1,0]	σουπερ	[4,0]	ευγενικ
εξαιρετικός	[1,0]	σκεπτικό	[1,0]	ομορφο	[1,0]	αποτελεσματική	[2,0]	θετικό
συγχρονος	[1,0]	ευχρηστο	[8,0]	καταλλήλο	[1,0]	ακριβώς	[6,0]	καταπλη
τιμίο	[1,0]	τιμία	[3,0]	πλούσια	[1,0]	πλούσιο	[1,0]	ευχρηστ
ωραίο	[4,0]	φορητός	[1,0]	γραφική	[1,0]	χρησιμη	[1,0]	οικονομ
ανταγωνιστικό	[-1,0]	εντυπωσιακό	[1,0]	τρομερο	[1,0]	διακριτικό	[1,0]	ποιότη
συγχρονη	[1,0]	υπεροχο	[1,0]					

Λίστα Αρνητικά Εξορυγμένων Λέξεων Γνώμης και Πολικότητα (Negative Filter Seed & Polarity)

Σύνολο: 32 λέξεις

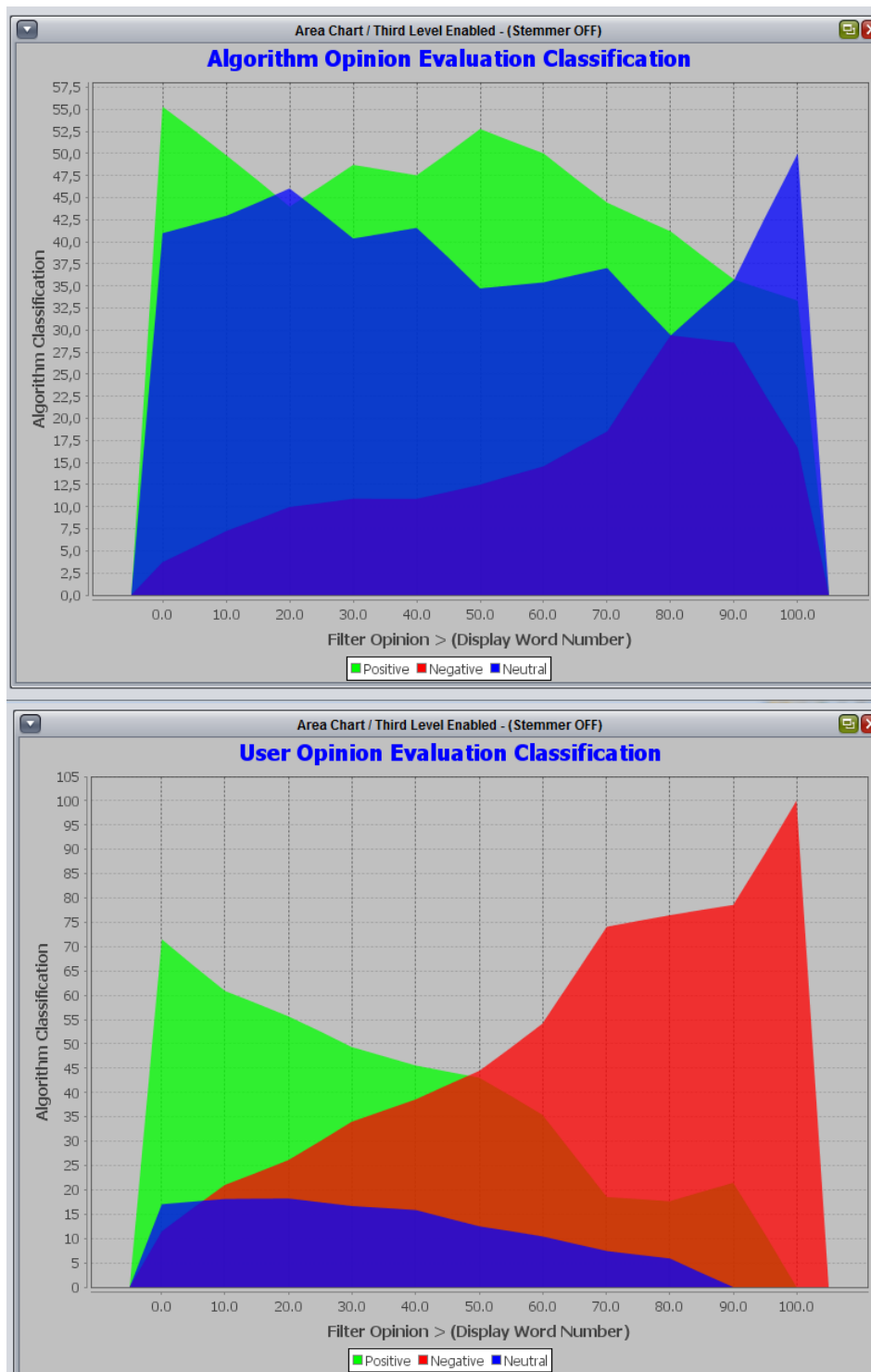


Εικόνα 11: Λίστα θετικών και αρνητικών λέξεων. Ταξινόμηση με βάση τη πολικότητα.



Εικόνα 12: Ανάλυση δεδομένων από τα manual δεδομένα

Στην συνέχεια το NiosTo πραγματοποιεί ανάλυση τρίτου επιπέδου δίνοντας μας με γραφικές παραστάσεις την ανάλυση του αλγόριθμου ταξινόμησης με τα φίλτρα εξόρυξης γνώμης.



Εικόνα 13: Ταξινόμηση σχολίων με βάση τα δεδομένα του χρήστη

Στο πάνω μέρος της εικόνας 13 παρατηρούμε την κατανομή των αποφάσεων του αλγόριθμου για διαφορετικό ποσοστό word number στην παράμετρο filter ενώ στο κάτω μέρος της εικόνας 13 την κατανομή για τις αποφάσεις που έχουν γίνει από τον

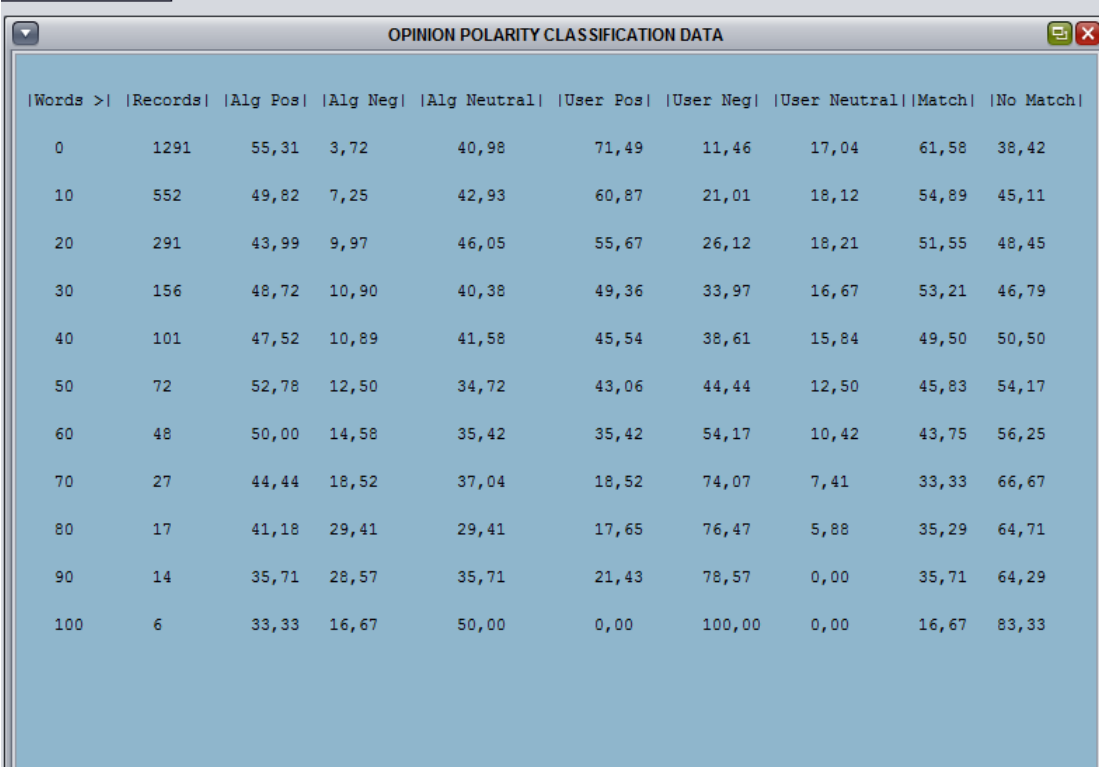
αξιολογητή. Παρατηρούμε πως δεν υπάρχει κάποια περιοχή στις τιμές του word number που οι τρεις τύποι αποφάσεων (αρνητικό, θετικό, ουδέτερο) να παραμένουν σταθεροί οπότε αυτό έχει αντίκτυπο και στο accuracy το οποίο (μέσο accuracy για όλες τις κλάσεις) παίρνει τιμές από 16 μέχρι 61 (βλ. Εικόνα 14).

Με τη γραφική παράσταση στην εικόνα 13 συμπεραίνουμε πως: *Accuracy*

Positive: 0.42-0.71

Negative: 0.35-0.90

Neutral: 0.10-0.20



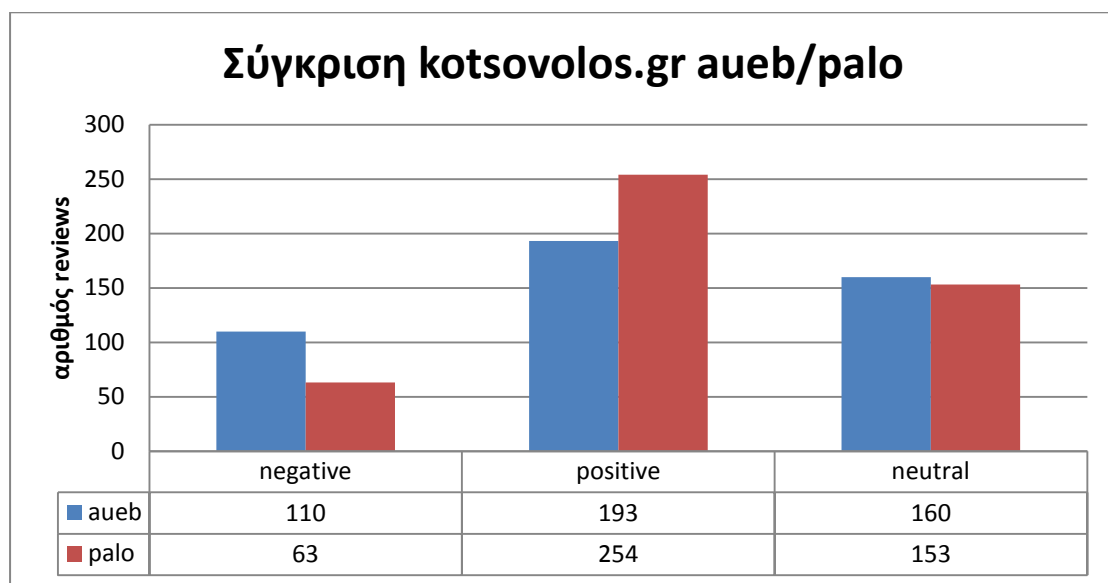
Words >	Records	Alg Pos	Alg Neg	Alg Neutral	User Pos	User Neg	User Neutral	Match	No Match
0	1291	55,31	3,72	40,98	71,49	11,46	17,04	61,58	38,42
10	552	49,82	7,25	42,93	60,87	21,01	18,12	54,89	45,11
20	291	43,99	9,97	46,05	55,67	26,12	18,21	51,55	48,45
30	156	48,72	10,90	40,38	49,36	33,97	16,67	53,21	46,79
40	101	47,52	10,89	41,58	45,54	38,61	15,84	49,50	50,50
50	72	52,78	12,50	34,72	43,06	44,44	12,50	45,83	54,17
60	48	50,00	14,58	35,42	35,42	54,17	10,42	43,75	56,25
70	27	44,44	18,52	37,04	18,52	74,07	7,41	33,33	66,67
80	17	41,18	29,41	29,41	17,65	76,47	5,88	35,29	64,71
90	14	35,71	28,57	35,71	21,43	78,57	0,00	35,71	64,29
100	6	33,33	16,67	50,00	0,00	100,00	0,00	16,67	83,33

Εικόνα 14: Γραφική παράσταση αλγόριθμου ταξινόμησης

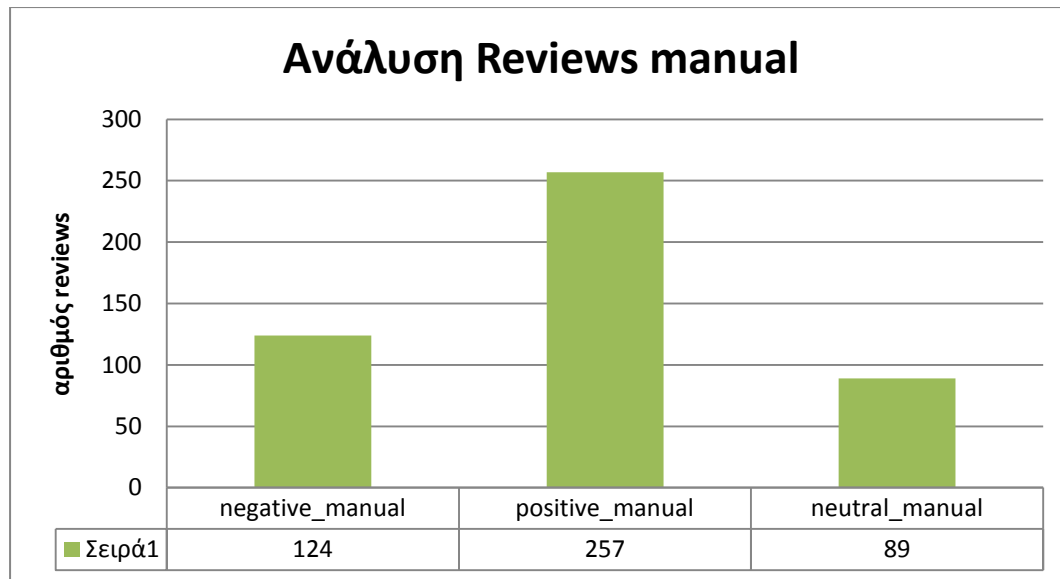
Στην Εικόνα 14 παρατηρούμε πως η στήλη match το Nioστο για τα δεδομένα του παραδείγματος public.gr μας εμφανίζει διακύμανση στο accuracy το οποίο αμέσως πέφτει από το 61,58 σε τιμές γύρω στο 50.

4.3 Αξιολόγηση των υπολοίπων εφαρμογών

Παρακάτω έχουμε τη σύγκριση των σχολίων πελατών από το ηλεκτρονικό κατάστημα kotsonolos.gr με 470 σχόλια κειμένου που υπάρχουν στο site της εταιρίας. Τα σχόλια αυτά είναι περασμένα σε αρχείο excel και υπάρχουν τα αποτελέσματα των δύο προγραμμάτων sentiment analysis που χρησιμοποίησα για κάθε σχόλιο (AUEB, Palo Pro). Για να μπορέσουμε να ελέγξουμε την απόδοση των συστημάτων και την ακρίβειά τους θεώρησα απαραίτητα να υπάρξει πίνακας με τις πραγματικές τιμές των σχολίων (Θετικό, αρνητικό ουδέτερο) τις οποίες μπορείτε να δείτε στον πίνακα παρακάτω.



Εικόνα 15: Πίνακας σύγκρισης αποτελεσμάτων προγραμμάτων aueb - Palo Pro



Εικόνα 16: Πραγματική ανάλυση με βάση τη πολικότητα

Για να μπορέσουμε να αξιολογήσουμε το μοντέλο μας χρησιμοποιούμε ένα πίνακα σύγκρισης (confusion matrix).

	Πρόβλεψη (Predicted Class)		
Πραγματική (Actual Class)		Class=Yes	Class=No
	Class= Yes	F ₁₁	F ₁₀
	Class=No	F ₀₁	F ₀₀

Όπου έχουμε:

F₁₁: TP, True positive

F₁₀: FN, False negative

F₀₁: FP, False positive

F₀₀: TN, True negative

$$p(\text{precision}) = \frac{TP}{TP + FP}$$

$$r(\text{recall}) = \frac{TP}{TP + FN}$$

Confusion Matrix: kotsovolos.gr aueb

Πρόβλεψη

πραγματική		pos	neu	neg	
	pos	142	61	53	256
	neu	21	57	7	85
	neg	30	42	50	122
		193	160	110	

Στο παρακάτω πίνακα μετράμε το σταθμισμένο μέσο όρο για ακρίβεια & ευαισθησία:

Kotsovolos_AUEB_Ακρίβεια:

Precision	0,515516
-----------	----------

Kotsovolos_AUEB_Ευαισθησία:

Recall	0,545037
--------	----------

Αντίστοιχα αν υπολογίσουμε το overall accuracy για τις 3 κλάσεις (όλες οι σωστές εκτιμήσεις προς το σύνολο των σχολίων) προκύπτει 0.53

Confusion Matrix kotsovolos.gr palo

Πρόβλεψη

πραγματική		pos	neu	neg	
	pos	217	32	8	257
	neu	13	72	4	89
	neg	24	49	50	123
		254	153	63	

Στον παρακάτω πίνακα μετράμε το σταθμισμένο μέσο όρο για ακρίβεια & ευαισθησία:

Kotsonolos_Palo_Ακρίβεια:

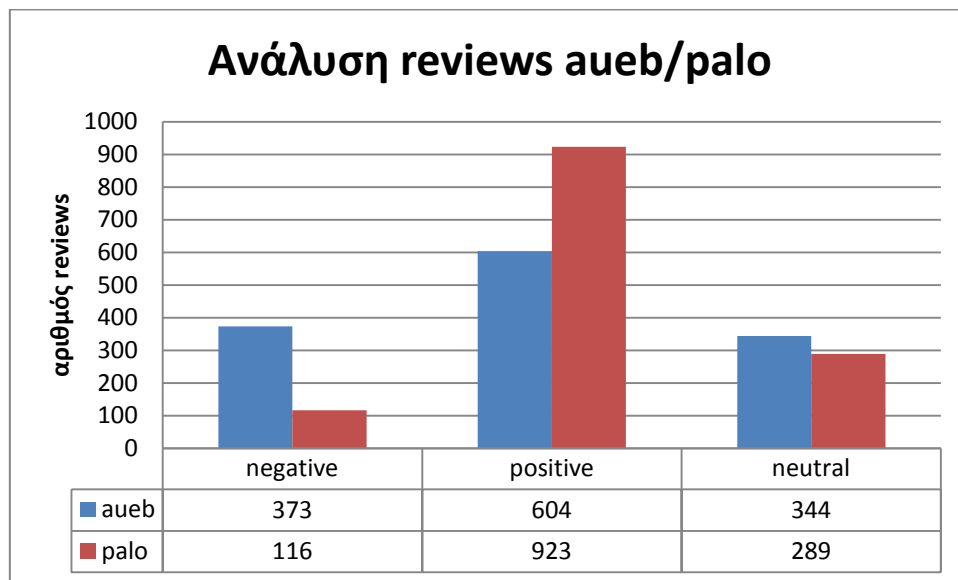
Precision	0,710457
-----------	----------

Kotsonolos_Palo_Ευαισθησία:

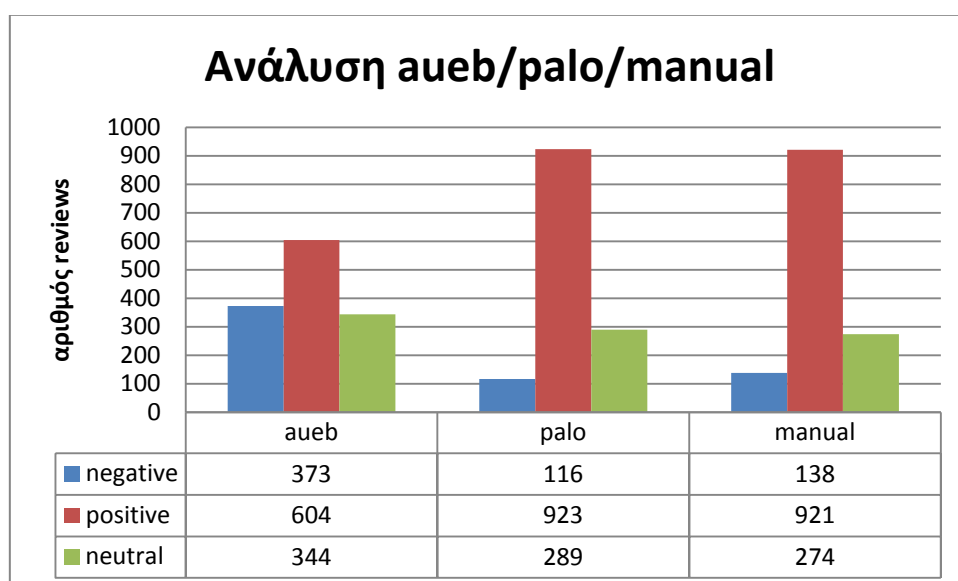
Recall	0,686617
--------	----------

Αντίστοιχα αν υπολογίσουμε το overall accuracy για τις 3 κλάσεις (όλες οι σωστές εκτιμήσεις προς το σύνολο των σχολίων) προκύπτει 0.72

Παρακάτω έχουμε τη σύγκριση των σχολίων πελατών από το ηλεκτρονικό κατάστημα public.gr με 1333 σχόλια κειμένου που υπάρχουν στο site της εταιρίας. Τα σχόλια αυτά είναι περασμένα σε αρχείο excel και υπάρχουν τα αποτελέσματα των δύο προγραμμάτων sentiment analysis που χρησιμοποίησα για κάθε σχόλιο (AUEB, Palo Pro). Για να μπορέσουμε να ελέγξουμε την απόδοση των συστημάτων και την ακρίβειά τους θεώρησα απαραίτητα να υπάρχει πίνακας με τις πραγματικές τιμές των σχολίων (Θετικό, αρνητικό ουδέτερο) τις οποίες μπορείτε να δείτε στον πίνακα παρακάτω.



Εικόνα 17: Πίνακας σύγκρισης αποτελεσμάτων προγραμμάτων aueb - Palo Pro



Εικόνα 18: Συγκριτικός πίνακας δύο συστημάτων με τις πραγματικές τιμές

Για να μπορέσουμε να αξιολογήσουμε το μοντέλο μας χρησιμοποιούμε ένα πίνακα σύγχυσης (confusion matrix).

Confusion Matrix public.gr aueb

Πρόβλεψη

	Πρόβλεψη			
	pos	neu	neg	
	pos	543	159	215
	neu	34	164	68
	neg	27	21	90
πραγματική		604	344	373

Στο παρακάτω πίνακα μετράμε το σταθμισμένο μέσο όρο για ακρίβεια & ευαισθησία:

Public_AUEB_Ακρίβεια:	Precision	0,539013
------------------------------	------------------	-----------------

Public_AUEB_Ευαισθησία:	Recall	0,620288
--------------------------------	---------------	-----------------

Αντίστοιχα αν υπολογίσουμε το overall accuracy για τις 3 κλάσεις (όλες οι σωστές εκτιμήσεις προς το σύνολο των σχολίων) προκύπτει 0.59

Confusion Matrix public.gr palo

Πρόβλεψη

	Πρόβλεψη			
	pos	neu	neg	
	pos	886	22	11
	neu	16	252	3
	neg	21	15	102
πραγματική		923	289	116

Στο παρακάτω πίνακα μετράμε το σταθμισμένο μέσο όρο για ακρίβεια & ευαισθησία:

Public_Palo_Ακρίβεια:	Precision	0,903732
Public_Palo_Ευαισθησία:	Recall	0,877704

Αντίστοιχα αν υπολογίσουμε το overall accuracy για τις 3 κλάσεις (όλες οι σωστές εκτιμήσεις προς το σύνολο των σχολίων) προκύπτει 0.93

5 Αξιολόγηση σχολίων από την πλευρά των επιχειρήσεων

5.1 Online Customer Reviews

Τα τελευταία χρόνια ολοένα και αυξάνονται οι ηλεκτρονικές αγορές σε διάφορα ηλεκτρονικά καταστήματα. Όταν αγοράζει ένας πελάτης από ηλεκτρονικό κατάστημα έχει άμεσα τη δυνατότητα να γράψει το σχόλιό του για το προϊόν, την εταιρία και γενικότερα τι του άρεσε και τι όχι στην συναλλαγή που πραγματοποίησε. Τα online σχόλια των πελατών στα website των εταιριών προσφέρουν στις ίδιες τις εταιρίες μια «γεύση» από την αντίδραση των πελατών και την ικανοποίηση ή όχι που δέχονται. Θα λέγαμε ότι παίζουν ένα σημαντικό ρόλο στην επικοινωνία μάρκετινγκ [28]

Οι πελάτες στα ηλεκτρονικά καταστήματα όλο και περισσότερο βασίζονται σε περιεχόμενο άλλων πελατών των ίδιων προϊόντων παρά στα σχόλια που γράφει η εταιρία για το προϊόν. Υπάρχουν πολλές έρευνες οι οποίες δείχνουν ότι το συναίσθημα των σχολίων των αγοραστών επηρεάζει σημαντικά τις πωλήσεις. Για παράδειγμα οι Chevalier και Mayzlin (2006) με την έρευνά τους μας δηλώνουν πως οι αξιολογήσεις των χρηστών στο διαδίκτυο επηρεάζουν σημαντικά τις πωλήσεις στην αγορά του βιβλίου. Από την άλλη πλευρά πολλές προσπάθειες έχουν γίνει για την ανάπτυξη νέων μεθόδων που θα παρέχουν μια συνοπτική επισκόπηση των σχολίων των πελατών με σκοπό να αποφευχθεί η υπερφόρτωση των πληροφοριών [29]

5.2 Online advertising

Η online διαφήμιση έχει κερδίσει την αποδοχή και το μερίδιο αγοράς εξαιτίας της καλύτερης στόχευσης στο κοινό της. Σύμφωνα με την IAB έκθεση [30] για το 2014 τα ετήσια έσοδα από τις online διαφημίσεις ανήλθαν σε 49.5 δισεκατομμύρια δολάρια, 15,6% υψηλότερο από το 2013. Στη σχετική έρευνα μάρκετινγκ αναφέρεται ότι η online διαφήμιση έχει καθιερωθεί ως ένα βασικό μέτρο μάρκετινγκ των επιχειρήσεων καθώς ενισχύει την πρόθεση αγοράς του πελάτη όταν το περιεχόμενο της διαφήμισης είναι σχετικό με την ιστοσελίδα που παρουσιάζεται. Οι Tam και Ho (2006) πρότειναν πως σχετικό περιεχόμενο στις διαφημίσεις κάθε ιστοσελίδας ενισχύει τη γνωστική απόδοση και βοηθάει τις ιστοσελίδες να αποκτήσουν ανταγωνιστικό πλεονέκτημα.

5.3 Πως οι καταναλωτές κρίνουν τη ποιότητα (quality) ενός eshop

Τα online καταστήματα λιανικής πώλησης διακρίνονται για την αλληλεπίδραση των πελατών (customer interaction) με την ιστοσελίδα [31], την παράδοση του προϊόντος και την ικανότητα να αντιμετωπίσει τα προβλήματα όταν αυτά δημιουργηθούν. Η ικανότητα των πελατών να πραγματοποιούν άμεσα 24 ώρες

την ημέρα συναλλαγές έχει αλλάξει τον τρόπο διαχείρισης των πελατειακών σχέσεων κάθε επιχείρησης. Το ηλεκτρονικό εμπόριο έχει εξαλείψει δύο από τα σημαντικότερα προβλήματα στην παροχή ποιότητας σε ένα περιβάλλον λιανικής πώλησης. Αρχικά έχει ελαχιστοποιηθεί η «ετερογένεια», παρέχοντας μια πολύ πιο συνεπή εμπειρία σε κάθε πελάτη. Αν συγκρίνουμε δηλαδή ένα υπάλληλο ενός καταστήματος με το website της εταιρίας θα λέγαμε ότι η ιστοσελίδα δε φτάνει ποτέ αργοπορημένη στη δουλειά, δεν είναι ποτέ σε κακή διάθεση και απρόσεκτη. Ένα ηλεκτρονικό κατάστημα δε ξεχνάει να σου προτείνει συμπληρωματικά προϊόντα με το κύριο προϊόν που αγοράζεις όπως επίσης βλέπει και τις προηγούμενες αγορές του ίδιου πελάτη και προτείνει συναφή προϊόντα.

Το ηλεκτρονικό εμπόριο έχει σαφώς αρκετά πλεονεκτήματα σε σχέση με το λιανικό εμπόριο αλλά πως μπορεί το ένα ηλεκτρονικό κατάστημα να ξεχωρίσει από το άλλο; Έρευνες έχουν δείξει ότι τα ηλεκτρονικά καταστήματα «πέφτουν» στη παγίδα ενός ατελείωτου «πολέμου» τιμών ενώ οι καταναλωτές και χρήστες του διαδικτύου είναι έτοιμοι να πληρώσουν ακριβά ένα προϊόν όταν το ηλεκτρονικό κατάστημα έχει καλύτερη φήμη.

Το πώς οι καταναλωτές όμως κρίνουν τη ποιότητα ενός ηλεκτρονικού καταστήματος ερεύνησαν οι Joel E. Collier and Carol C. Bienstock (MIT, 2006). Η έρευνα πραγματοποιήθηκε σε μεγάλη πόλη και οι ερωτηθέντες είχαν αγοράσει προϊόντα από ηλεκτρονικό κατάστημα. Οι τέσσερις κατηγορίες ερωτήσεων τις έρευνας αναλύονται παρακάτω.

Η πρώτη κατηγορία ερωτήσεων εστιάζοταν στην αλληλεπίδραση του πελάτη με την ιστοσελίδα. Ερωτήσεις σχετικά με την ευκολία χρήσης, την πλοήγηση, την προστασία της ιδιωτικής ζωής, το σχεδιασμό του website, τα γραφικά, τα πιθανά τεχνικά προβλήματα του site, την ακρίβεια των πληροφοριών και την αντικειμενικότητά τους.

Η δεύτερη κατηγορία ερωτήσεων ζήτησε από τους πελάτες να αξιολογήσουν τα ηλεκτρονικά καταστήματα με βάση την παράδοση του προϊόντος που αγόρασαν. Οι ερωτήσεις είχαν να κάνουν με την ακρίβεια, την φροντίδα που είχαν συσκευαστεί τα προϊόντα και την κατάσταση παραλαβής του προϊόντος.

Στο τρίτο μέρος των ερωτήσεων ρωτήθηκαν οι συμμετέχοντες αν δημιουργήθηκε κάποιο πρόβλημα στη συναλλαγή και πόσο γρήγορα το ηλεκτρονικό κατάστημα ανταποκρίθηκε και έλυσε το πρόβλημα. Ερωτήσεις σχετικές με την ανησυχία της εταιρείας και την ικανότητά της να επιλύσει το ζήτημα ήταν σε αυτή τη κατηγορία.

Τέλος στο τέταρτο μέρος της έρευνας οι ερωτήσεις απευθύνονται στη συνολική ικανοποίηση του πελάτη από τα ηλεκτρονικά καταστήματα και αν είναι διατεθειμένος να ακολουθεί τα ηλεκτρονικά καταστήματα και στο μέλλον.

Ποια είναι τα χαρακτηριστικά εκείνα τα οποία κάνουν ένα χρήστη ενός ηλεκτρονικού καταστήματος να γίνει πελάτης ;

Αλληλεπίδραση (interaction) με την ιστοσελίδα.

Η ιστοσελίδα μια εταιρείας λιανικής πώλησης είναι το πρωταρχικό σημείο επαφής σε μια ηλεκτρονική συναλλαγή. Εκεί οι πελάτες βρίσκουν πληροφορίες σχετικά με την εταιρεία και τα προϊόντα ενώ η εταιρεία αποσκοπεί στην ικανοποίηση του πελάτη σε κάθε στάδιο της διαδικασίας αγοράς (έρευνα, αγορά, after sales). Εάν μια ιστοσελίδα είναι δύσκολη στη χρήση ή γραφιστικά «δύσκολη» ώστε να τη χρησιμοποιήσει ο μέσος καταναλωτής τότε οι υποψήφιοι πελάτες θα κατευθυνθούν σε άλλες ιστοσελίδες. Υπάρχουν πέντε χαρακτηριστικά τα οποία επηρεάζουν το πώς οι πελάτες αξιολογούν την ποιότητα των αλληλεπιδράσεων στη ιστοσελίδα της εταιρείας και αυτά είναι τα εξής:

1. Ευκολία στη χρήση: όταν οι πελάτες έχουν πρόσβαση στην ηλεκτρονική σελίδα μιας εταιρείας λιανικής συχνά διαπιστώνουν πως δεν παρέχεται καμία επίσημη οδηγία για το πώς θα ολοκληρώσουν μια συναλλαγή. Αυτό σημαίνει ότι μια τοποθεσία στο διαδίκτυο πρέπει να είναι εξαιρετικά εύκολη και απλή ανεξάρτητα από το πόση εμπειρία έχει ένας πελάτης με τους ηλεκτρονικούς υπολογιστές. Οι πελάτες ζητούν εύκολες μηχανές αναζήτησης μέσα στο ηλεκτρονικό κατάστημα ώστε να βρίσκουν εύκολα και γρήγορα τα προϊόντα τους. Το βασικό πλεονέκτημα του ηλεκτρονικού εμπορίου είναι ο ατελείωτος χώρος για ράφια όπου μπορεί άμεσα να μετατραπεί και σε σημαντικό πρόβλημα αν η ιστοσελίδα είναι δύσκολη στη χρήση.
2. Ασφάλεια: Οι πελάτες των ηλεκτρονικών καταστημάτων θέλουν να είναι σίγουροι πως οι συναλλαγές τους και οι πληροφορίες που ανταλλάσσουν με την εταιρεία είναι απόλυτα ασφαλείς. Για παράδειγμα οι πελάτες πρέπει να αισθάνονται ότι τα στοιχεία της πιστωτικής τους κάρτας είναι ασφαλή και προστατευμένα από διάφορους χάκερ. Τα πιστοποιητικά ασφαλείας των ηλεκτρονικών συναλλαγών που παρέχονται από διάφορους οργανισμούς είναι ένας λόγος για να αγοράσει κάποιος καταναλωτής από το συγκεκριμένο ηλεκτρονικό κατάστημα. Εκτός από τη πιθανότητα να «κλαπούν» στοιχεία της πιστωτικής κάρτας του πελάτη ένας άλλος κίνδυνος είναι η παραβίαση της ιδιωτικής ζωής του κάθε πελάτη. Δεν υπάρχει πιο γρήγορος τρόπος ένας ηλεκτρονικός λιανοπωλητής να αυξήσει την οργή των πελατών του από το να πουλήσει την ηλεκτρονική διεύθυνση, τον αριθμό του κινητού και άλλα προσωπικά δεδομένα σε τρίτες εταιρείες.
3. Απλός Σχεδιασμός: Ο απλός σχεδιασμός της ιστοσελίδας, το κείμενο, η μορφή, τα χρώματα είναι εξαιρετικά σημαντικά για τη διευκόλυνση μιας ηλεκτρονικής συναλλαγής. Θα μπορούσαμε να πούμε ότι είναι κάτι ανάλογο με την ατμόσφαιρα που υπάρχει μέσα σε ένα φυσικό κατάστημα της εταιρείας. Σε ένα ηλεκτρονικό κατάστημα θα πρέπει να προσέχουν και να αποφεύγουν τα πολλά animation καθώς και τα κόκκινα γράμματα σε μπλε φόντο διότι είναι δύσκολο να διαβαστούν.

4. Συνοχή και Ευελιξία: Για να χτίσει κάποιο ηλεκτρονικό κατάστημα μια πελατειακή βάση θα πρέπει να το διακρίνει η αξιοπιστία και η συνέπεια. Όταν υπάρχουν χρήστες της σελίδας που χρησιμοποιούν εκφράσεις όπως «το ηλεκτρονικό κατάστημα είναι συνέχεια κάτω» ή «δε δουλεύει ποτέ το site», τότε υπάρχει ζήτημα αξιοπιστίας της συγκεκριμένης ιστοσελίδας. Οι ηλεκτρονικοί έμποροι λιανικής θα πρέπει να γνωρίζουν πως όταν έχουν μια τοποθεσία web η οποία αντιμετωπίζει προβλήματα λειτουργικά είναι σα να υπάρχει μια ταμπέλα σε φυσικό κατάστημα και να αναγράφει «επιστρέφω σε λίγο». Οι υπερ-συνδέσεις δε θα πρέπει να οδηγούν σε αδιέξοδα ή να μην οδηγούν το πελάτη στην επιθυμητή ιστοσελίδα. Για να μπορεί να προσελκύσει ένα ηλεκτρονικό κατάστημα μεγάλο μερίδιο πελατών θα πρέπει να έχει για παράδειγμα πολυγλωσσικές μεταφράσεις της σελίδας.
5. Καλή πληροφορία: Οι μανατζερ του ηλεκτρονικού καταστήματος θα πρέπει να ελέγχουν και να διατηρούν ακριβείς πληροφορίες και ενημερωμένες. Αν στα προϊόντα που εμπορεύεται ένα ηλεκτρονικό κατάστημα δεν έχουν τα σωστά τεχνικά χαρακτηριστικά, τις σωστές φωτογραφίες, τα σωστά στοιχεία για τη παράδοση τους τότε είναι πολύ εύκολο να αποκτήσει δυσαρεστημένους πελάτες. Η λάθος τιμή που μπορεί να υπάρχει σε ένα online περιβάλλον επιδεινώνει την ενδεχόμενη άσχημη εντύπωση του πελάτη και οι πελάτες τελικά αισθάνονται απογοητευμένοι.

5.4 Τι επηρεάζει τα σχόλια των online πελατών

Τα social media και τα forum έχουν δημιουργήσει τα τελευταία χρόνια ένα τεράστιο όγκο συνομιλίας και σχολίων από στόμα σε στόμα γύρω από τις μάρκες και τα προϊόντα. Από την άποψη των επιχειρήσεων, τα social media και τα σχόλια των χρηστών σε eshop μπορούν να παρέχουν χρήσιμες πληροφορίες των πελατών για τα προϊόντα τους. Η επιθυμία να «ακούν» οι εταιρίες τα σχόλια των πελατών για τα προϊόντα είναι διάχυτη και για αυτό σαρώνουν τακτικά τα σχόλια αυτά ώστε να μετρήσουν το συναίσθημα του πελάτη (sentiment analysis). Για να δούμε όμως πως οι συνομιλίες αυτές αντιπροσωπεύουν την πραγματική υποκειμενική ψυχολογία των πελατών ενός προϊόντος χρειάζεται ανάλυση των σχολίων ώστε να γίνει διακριτή η συμπεριφορά του καταναλωτή.

5.5 Ποιος γράφει σχόλια και γιατί;

Ενώ οι πελάτες σε ένα ηλεκτρονικό κατάστημα αναζητούν τακτικά τις απόψεις άλλων πριν πραγματοποιήσουν κάποια συναλλαγή είναι πιο σπάνιο να γράφουν οι ίδιοι κάποιο σχόλιο (θετικό ή αρνητικό) για το προϊόν που αγόρασαν. Αυτό έχει σαν αποτέλεσμα οι απόψεις των δυνητικών αγοραστών και χρηστών των

social media να προέρχονται από ένα μικρό μόνο τμήμα του πληθυσμού το οποίο δεν αντικατοπτρίζει το αίσθημα όλων των πελατών.

5.6 Πως χρησιμοποιούν οι επιχειρήσεις τα σχόλια των πελατών:

1. Χάραξη στρατηγικής για τιμές ενάντια στον ανταγωνισμό (web offers). Όλα τα ηλεκτρονικά καταστήματα ακολουθούν τις τάσεις του ανταγωνισμού. Ένας υποψήφιος πελάτης είναι πολύ εύκολο να ενημερωθεί και να συγκρίνει τιμές προϊόντων και υπηρεσιών μέσα από το διαδίκτυο. Ο καταναλωτής δεν αποφασίζει μόνο από την τιμή, υπάρχουν και άλλοι παράγοντες. Οι επιχειρήσεις έχοντας λοιπόν τα σχόλια των χρηστών από τις online αγορές τους μπορούν να χαράξουν τη στρατηγική τους έναντι του ανταγωνισμού με βάση τη τιμή των προϊόντων τους. Έτσι πραγματοποιούν διάφορες προωθητικές ενέργειες και προσφορές μόνο για τα ηλεκτρονικά τους καταστήματα.
2. Διαχείριση εικόνας και διαφύλαξη καλής εικόνας του brand: Με τα σχόλια που λαμβάνουν οι επιχειρήσεις μέσα από email, social media, το ηλεκτρονικό τους κατάστημα βρίσκουν τα θετικά και τα αρνητικά τους σημεία από τη ματιά των πελατών τους. Για παράδειγμα αν μεγάλο ποσοστό καταναλωτών εκφράζει τη δυσαρέσκεια του για μια συμπεριφορά της εταιρίας ή μια πολιτική που ακολουθεί τότε οι υπεύθυνοι της επιχείρησης πραγματοποιούν διάφορες ενέργειες ώστε να διορθώσουν αυτή τη κακή εικόνα που έχει δημιουργηθεί.
3. Εύρεση λαθών ή ελαττωμάτων: Με τη βοήθεια των σχολίων των πελατών από τις αγορές που πραγματοποιούν στα ηλεκτρονικά καταστήματα οι εταιρίες δέχονται την απαραίτητη πληροφόρηση για τυχόν λάθη ή ελαττώματα σε διαδικασίες και προϊόντα ώστε να δημιουργήσουν νέες διαδικασίες ή τη βελτίωση των υπαρχόντων. Για παράδειγμα αν μεγάλο ποσοστό των πελατών ενός ηλεκτρονικού καταστήματος έχει πρόβλημα με την εταιρία μεταφορών που κάνει τις παραδόσεις των προϊόντων τότε η επιχείρηση θα πρέπει να το ερευνήσει και να προβεί στις απαραίτητες ενέργειες.
4. Χάραξη Στρατηγικής – Trends εποχής: Οι επιχειρήσεις κατά περιόδους αλλάζουν την στρατηγική που ακολουθούν. Σε αυτή την αλλαγή της στρατηγικής τους επηρεάζονται από τα σχόλια που γράφουν οι πελάτες τους, σε social media και το ηλεκτρονικό κατάστημά τους, αλλά και από την αλλαγή των τάσεων (trends) της εποχής. Για παράδειγμα αν μια εταιρία δέχεται θετικά σχόλια για να κάνει ένα άνοιγμα σε νέα αγορά τότε είναι ένα θετικό σημάδι για να πραγματοποιήσει αυτή την ανάπτυξη της επιχείρησης.
5. Δημιουργία προϊόντος – καμπάνιας: Με τη σωστή ανάλυση και αξιοποίηση των σχολίων των πελατών οι επιχειρήσεις καταφέρνουν να ενημερωθούν άμεσα από τους πελάτες τους για τις ανάγκες τους. Αυτό έχει σαν αποτέλεσμα

να είναι σε θέση να δημιουργήσουν κάποιο νέο προϊόν που θα καλύπτει τις ανάγκες των πελατών τους ή τη δημιουργία ανάλογης καμπάνιας.

6. Δημιουργία μακροχρόνιων σχέσεων: Η δημιουργία ενός CRM συστήματος που να βασίζεται στο προφίλ του πελάτη ανάλογα με τα σχόλιά του και το ιστορικό των αγορών του βοηθάει τις επιχειρήσεις να δημιουργήσουν μακροχρόνιες και πιο προσωποποιημένες σχέσεις με τους πελάτες τους.
7. Πιστότητα Πελατών: Ο ανταγωνισμός σε όλους τους κλάδους είναι μεγάλος και το ίδιο είναι και στα ηλεκτρονικά καταστήματα. Η κατάλληλη ανάλυση των σχολίων των πελατών σε συνδυασμό με τις απαραίτητες ενέργειες από πλευράς της επιχείρησης μπορούν να δημιουργήσουν αύξηση της επισκεψιμότητας του site, αύξηση των πωλήσεων και πιστούς πελάτες. Εφόσον ένα ηλεκτρονικό κατάστημα έχει ικανοποιημένους πελάτες αυξάνει την πιστότητα των πελατών του.

6 Μελέτες περίπτωσης

6.1 Case study Public

Για τη παρακάτω μελέτη πραγματοποιήσα συνέντευξη με την υπεύθυνη δημοσίων σχέσεων της εταιρίας και μας παρουσίασε τα παρακάτω. Τα Public είναι μια από τις μεγαλύτερες εταιρίες λιανικής πώλησης στην Ελλάδα. Είναι Ελληνική εταιρία με 48 καταστήματα σε Ελλάδα και Κύπρο, 2 online καταστήματα, 1 online κατάστημα για εισιτήρια και 1 online κατάστημα για βιβλία amazon. Διαθέτει ένα εταιρικό Blog και 5 κανάλια στα Social Media. Η εταιρία απασχολεί περίπου 2000 εργαζομένους, έχει 6000 προμηθευτές και περίπου 20εκ επισκέπτες ετησίως.

6.1.1 Η προτεραιότητά του Public

Καθημερινά η Public δέχεται εκατοντάδες αιτήματα από τους πελάτες της στα κανάλια επικοινωνίας που διαθέτει. Για παράδειγμα το 2014 δέχθηκε περίπου 562.000 interactions από τα οποία μόνο τα 650 είναι καταγεγραμμένα παράπονα. Αυτό σημαίνει πως σαν εταιρία έχει προτεραιότητα τις ανάγκες του πελάτη επεμβαίνοντας προτού η ανάγκη γίνει παράπονο.

6.1.2 Η ποικιλία των σχολίων των πελατών

Τα σχόλια και οι παρατηρήσεις που δέχεται καθημερινά η εταιρία έχουν να κάνουν με:

- Εξυπηρέτηση
- Καθυστερήση παραγγελίας
- Ανοιγμένη συσκευασία – Χρησιμοποιημένο προϊόν
- Φθαρμένο προϊόν
- Τιμή
- Λάθος Προϊόν, γκάμα, έξοδα αποστολής κτλ.

Τα παραπάνω είδη σχολίων των πελατών η εταιρία Public τα κατηγοριοποιεί ανάλογα με τη σοβαρότητα της κατάστασης, τη φύση του συγκεκριμένου αιτήματος και το κανάλι το οποίο προήλθε το αίτημα. Όσον αφορά τη σοβαρότητα της κατάστασης το κάθε αίτημα λαμβάνει χαρακτηρισμό low, medium, high, red, crisis. Τέλος σχετικά με τη φύση του κάθε αιτήματος τα σχόλια κατηγοριοποιούνται ως καθυστέρηση παραλαβής, βιογραφικό σημείωμα κτλ.

6.1.3 Οι βασικοί πυλώνες

Θα μπορούσαμε να πούμε πως η εταιρία στηρίζετε σε τρεις βασικούς πυλώνες ώστε να διαχειριστεί με αποτελεσματικότητα και ταχύτητα τα αιτήματα των πελατών της.

- Κουλτούρα : Για να επιτύχει κάποιος οργανισμός την αποτελεσματική διαχείριση των εκατοντάδων αιτημάτων που δέχεται καθημερινά από πελάτες του πρέπει να έχει την εξυπηρέτηση του πελάτη στο επίκεντρο της κουλτούρας του. Θα μπορούσαμε να πούμε πως την εξυπηρέτηση του πελάτη, η εταιρία Public, την πλαισιώνει η οργανωτική δομή της εταιρίας, το προσωπικό, η καθημερινή εξυπηρέτηση, το ομαδικό πνεύμα και τα αποτελέσματά της. Στο οργανόγραμμά της η εταιρία έχει πολλά διαφορετικά σημεία επαφής στη διάθεση των πελατών (καταστήματα, eshop, τηλεφωνικό κέντρο, live chat, social media) ούτως ώστε να έχει μια άμεση επικοινωνία με το πελάτη 24 ώρες το 24ωρο.
- Σύστημα Διαχείρισης παραπόνων: Η εταιρία Public διαθέτει ένα ελληνικό σύστημα διαχείρισης αιτημάτων φτιαγμένο εξολοκλήρου στα μέτρα της εταιρίας. Σχεδιάστηκε και δημιουργήθηκε με σκοπό να καλύψει τις ανάγκες της επιχείρησης ώστε να εξυπηρετεί το μοναδικό για τα ελληνικά δεδομένα Brand, το πάθος της εταιρίας για έγκαιρη και σωστή εξυπηρέτηση του πελάτη και τέλος τα υψηλά KPIs σε αυτό τον τομέα.
- Daily monitoring & Μέτρηση αποτελεσμάτων: Ο τρίτος βασικός πυλώνας στη διαχείριση των αιτημάτων των πελατών της Public είναι η καθημερινή παρακολούθηση των αιτημάτων στα eshop της, στα social media & στα forum. Τέλος μετράει τα αποτελέσματα και κάνει έρευνα για τυχόν βελτιώσεις των διαδικασιών.

CRM σύστημα

Τα PUBLIC διαθέτουν ένα ολοκληρωμένο σύστημα CRM για την διασφάλιση της παροχής βέλτιστης εξυπηρέτησης προς τους πελάτες. Προκειμένου να αυξήσουν την πιστότητα και συχνότητά των επισκέψεών τους και να τους εξυπηρετήσουν με τον καλύτερο δυνατό τρόπο, δημιούργησαν τα μεγαλύτερα club Τεχνολογίας & πολιτισμού, μέσω των οποίων τα μέλη απολαμβάνουν αποκλειστικά προνόμια. Παράλληλα επικοινωνούν μαζί τους μέσω διαφόρων καναλιών επικοινωνίας (Email/sms/directmail/Callcenter) της εφαρμογής του Microsoft Dynamics CRM για τη διεκπεραίωση των αιτημάτων τους και ενημέρωσή τους.

Μέσα από τις κάρτες MyPublic & KidsClub συγκεντρώνονται τα buyinghabits των πελατών, ώστε σε περίπτωση παραπόνου να συνδέουμε άμεσα τον πελάτη με το ιστορικό αγορών του και τις αγοραστικές του συνήθειες. Έτσι σε περίπτωση παραπόνου γίνεται άμεση έρευνα ώστε να έχουμε το ιστορικό του πελάτη και ακολουθείται άμεσα επίσης recovery process στο 80% των περιπτώσεων.

6.1.4 Διαδικασίας Προσέλκυσης & Επιλογής Προσωπικού εξασφαλίζει τη βέλτιστη Εξυπηρέτηση Πελατών

Ο μοναδικός χαρακτήρας του Public στηρίζεται στην **βαθιά γνώση των πολλών προϊόντικών κατηγοριών**(τεχνολογία, υπηρεσίες επικοινωνίας, βιβλίο – μουσική κλπ) αλλά και στο **κέφι - πάθος για εξυπηρέτηση** των επισκεπτών. Για να διασφαλιστούν αυτά ακολουθείται η εξής διαδικασία :

1. Επακριβής Προσδιορισμός αρμοδιοτήτων και καθηκόντων καθώς και προκαθορισμένα προσόντα και τεχνικές γνώσεις ανά θέση front line (πχ πωλητής βιβλίου, call center agent για τηλεπικοινωνίες κλπ)
2. Αξιολόγηση από δύο τουλάχιστον ανθρώπους με εμπειρία στην εξυπηρέτηση πελατών / εκπαίδευση πωλητών
3. Case studies σε ειδικές γνώσεις της κατηγορίας αλλά και soft skills εξυπηρέτησης
4. Δοκιμαστική περίοδος εργασίας στο front line κατά την οποία ο νέος εργαζόμενος αξιολογείται ως προς την απόδοσή του αλλά και την ικανότητά του να ενταχθεί στην εταιρεία

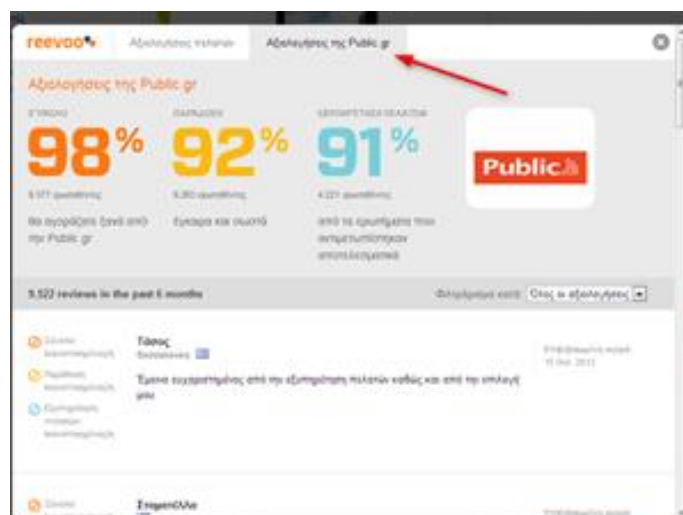
6.1.5 Διαχείριση των Αιτημάτων των πελατών της εταιρίας η οποία υποστηρίζει την Εξυπηρέτηση Πελατών

Οποιοδήποτε αίτημα, φτάνει στο κατάστημα ή στο site της εταιρίας απαντάται εντός 24ώρου. Για παράδειγμα με επίσκεψη στο κατάστημα ο πελάτης μπορεί να παραγγείλει οποιοδήποτε προϊόν επιθυμεί και δεν βρίσκει. Παραλαμβάνει έντυπο με τον κωδικό της παραγγελίας (web order id) και μπορεί να ενημερωθεί για την εξέλιξη της ανά πάσα στιγμή. Στη συνέχεια λαμβάνει ενημέρωση τελικής διαθεσιμότητας και χρόνου παραλαβής και μπορεί να παραλάβει το προϊόν του είτε από το κατάστημα της επιλογής του, είτε με αποστολή στο χώρο του.

Η εξυπηρέτηση και ικανοποίηση του πελάτη παραμένει πάντα στο κέντρο της φιλοσοφίας του Public και λαμβάνεται πάντα υπόψη στον καθορισμό των ατομικών και ομαδικών στόχων. Επιπλέον μέσα από τις συνέργειες που δημιουργούν (για παράδειγμα με το amazon) πετυχαίνει η εταιρία πιο γρήγορες και αποτελεσματικές διαδικασίες οι οποίες την ξεχωρίζουν από τον ανταγωνισμό και συμβάλουν και αυτές στην εξυπηρέτηση του πελάτη με τον καλύτερο δυνατό τρόπο. Όλα τα παραπάνω έχουν συντελέσει στις θετικές αξιολογήσεις των πελατών για την ομάδα του Public.

Η υπηρεσία Reenoo είναι μια αμερόληπτη, ανεξάρτητη και διαφανής εταιρεία συλλογής και διαμοιρασμού αξιολογήσεων, καταναλωτικών προϊόντων πάσης φύσεως.

Δεν επηρεάζεται από κατασκευαστές ή καταστήματα και δημοσιεύει όλες τις αξιολογήσεις, καλές ή κακές οι οποίες είναι γνήσιες και έχουν γραφτεί από άτομα που κατέχουν το προϊόν. Τα Public συνεργάζονται με την Reenoo και όπως αποδεικνύουν και τα συνημμένα αρχεία, τα scores που έχουμε επιτύχει στο Public.gr καθώς και τα αντίστοιχα του κύριου ανταγωνιστή, που επίσης συνεργάζεται με τη Reenoo, δείχνουν το συντριπτικό ποσοστό των θετικών αξιολογήσεων των πελατών τους έναντι του δεύτερου.



Εικόνα 19: Σχόλια πελατών από Reenoo για τον πελάτη public.gr

6.1.6 Παραδείγματα Αιτημάτων – Σχολίων πελατών

Κάθε αίτημα πελάτη επεξεργάζεται και στη συνέχεια αποτελεί κίνητρο για βελτιώσεις και αλλαγές στις διαδικασίες που ακολουθεί η εταιρία όπως επίσης για τη δημιουργία νέων διαδικασιών και μεθόδων.

Για παράδειγμα το αίτημα των πελατών, Ελλήνων του εξωτερικού, να παραγγέλνουν και να παραλαμβάνουν ελληνικά βιβλία στο εξωτερικό κινητοποίησε τα τμήματα commercial, IT, stores και webfulfillment και ενεργοποίησε τη συνεργασία των Public με το amazon κάνοντας δυνατή την αποστολή 50.000 ελληνικών προϊόντων πολιτισμού που διαθέτουμε σε όλες τις χώρες που δραστηριοποιείται το amazon.

Όταν ένα σχόλιο ενός πελάτη ήταν «Θέλω να μπορώ να αγοράζω όλα μου τα βιβλία/χαρτικά από ένα σημείο» τότε κινητοποιήθηκαν τα τμήματα Commercial, Stores, Web για να πετύχουν το καλύτερο δυνατό αποτέλεσμα ώστε το αίτημα αυτό να μην γίνει παράπονο. Το αποτέλεσμα ήρθε το Σεπτέμβριο 2013 όπου κατά τη περίοδο του Back2school υπήρξε η δυνατότητα αγοράς βιβλίων, ξενόγλωσσα, φροντιστηριακά, IB, ΟΕΔΒ, και χαρτικά online ή στο κατάστημα.

Και τέλος ένα παράδειγμα όπου το Public ανταποκρίθηκε αποτελεσματικά ήταν όταν ένας πελάτης δεν ήθελε να μείνει χωρίς κινητό τηλέφωνο όταν αυτό χάλασε. Το σχόλιο αυτό το ανάρτησε στο ηλεκτρονικό κατάστημα της εταιρίας και κινητοποιήθηκαν τα τμήματα Services, Stores, Customer Passion ώστε το Νοέμβριο

του 2013 να ξεκινήσει να υπάρχει σε τέτοιες περιπτώσεις επίσκεψη στο σπίτι από τεχνικό, back up και προσωρινή αντικατάσταση συσκευής.

6.2 Case study Κωτσόβολος

Για την παρακάτω μελέτη περίπτωσης πραγματοποιήσα συνέντευξη με το διευθυντή του Digital & Content Marketing της εταιρίας ο οποίος μας ενημερώνει για τα παρακάτω. Η Κωτσόβολος, η Νο1 αλυσίδα ηλεκτρικών και ηλεκτρονικών συσκευών στην Ελλάδα, κατέκτησε την 1η θέση στην έρευνα «Famous Brands» αποδεικνύοντας ότι βρίσκεται σταθερά στην κορυφή των προτιμήσεων των πελατών στην ελληνική αγορά.

Στην έρευνα που διεξήγαγε η Hill + Knowlton Strategies με τη μεθοδολογία του Reputation Institute, η εταιρεία βραβεύεται για 9η χρονιά στην κατηγορία «Καταστήματα Ηλεκτρονικών και Ηλεκτρικών Συσκευών». Η διάκριση αυτή είναι ιδιαίτερα σημαντική καθώς προέρχεται από αυθόρμητες απαντήσεις των πελατών, οι οποίοι ξεχωρίζουν την Κωτσόβολος ως την εταιρεία με την ισχυρότερη φήμη σε αυτή την κατηγορία.

Για την εταιρεία το γεγονός ότι αναδεικνύεται σταθερά στην 1η θέση των Famous Brands αποτελεί μεγάλη τιμή αλλά και απόδειξη της σχέσης εμπιστοσύνης που έχει αναπτύξει με τους πελάτες της καταφέροντας πάντα να πρωταγωνιστεί στις εξελίξεις.

6.2.1 Ηλεκτρονικό κατάστημα

Τα δυνατά σημεία στα οποία στηρίζετε το ηλεκτρονικό κατάστημα της Κωτσόβολος είναι:

- Απίστευτη ταχύτητα πλοήγησης
- Καλύτερη διαθεσιμότητα από ποτέ
- Ολοκληρωμένες προτάσεις & έξυπνες αγορές
- Collect @ Store
- Πλούσιο περιεχόμενο
- Πανεύκολη αναζήτηση

Το ηλεκτρονικό κατάστημα της Κωτσόβολος ενσωματώνει μια σειρά από καινοτομίες για να κάνει τις αγορές άνετες, εύκολες και πάντα με ασφάλεια.

Μερικά ακόμη από τα πρακτικά πλεονεκτήματα του ηλεκτρονικού καταστήματος της ΚΩΤΣΟΒΟΛΟΣ είναι:

- οι πολλαπλοί τρόποι πληρωμής όπως πιστωτική κάρτα, Πλάνο Δόσεων ΚΩΤΣΟΒΟΛΟΣ και αντικαταβολή σε όλη την Ελλάδα.
- οι πολλαπλοί τρόποι παράδοσης όπως παράδοση κατ' οίκον και collect @ store με παραλαβή από τα ιδιόκτητα φυσικά καταστήματα σε 70 σημεία σε όλη τη χώρα.
- το σύστημα αξιολόγησης REEVOO, όπου οι καταναλωτές μπορούν να δουν τις κριτικές άλλων χρηστών για το προϊόν που ενδιαφέρονται να αποκτήσουν.
- η σταθερά κορυφαία ασφάλεια συναλλαγών που εξασφαλίζεται με τη χρήση των τελευταίων διεθνών προτύπων και τη σφραγίδα εγγύηση των Norton Secured, McAfee Secure, Mastercard SecureCard και Verified by Visa.
- οι ηλεκτρονικές αγορές στο www.kotsovolos.gr περιμένουν όλους τους απαιτητικούς και ενημερωμένους καταναλωτές για να τους προσφέρουν την καλύτερη εξυπηρέτηση.

Η ΚΩΤΣΟΒΟΛΟΣ διακρίθηκε στην ενότητα Redesign-Relaunch, στη φετινή βράβευση των E-volution Awards 2014, παραλαμβάνοντας το βραβείο silver για το καλύτερο site που λανσαρίστηκε εκ νέου.

6.2.2 Η ποικιλία των σχολίων των πελατών

Η Κωτσόβολος εξυπηρετεί καθημερινά εκατοντάδες αιτήματα πελατών από τα βασικά κανάλια επικοινωνίας της με τους πελάτες. Τα κανάλια αυτά είναι το τηλεφωνικό κέντρο, το ηλεκτρονικό κατάστημα μέσα από τη καρτέλα επικοινωνία, από τα social media και μέσα από την αλυσίδα καταστημάτων της. Τα σχόλια αυτά κυρίως αφορούν:

- Προϊόντα – Υπηρεσίες
- Ηλεκτρονικές αγορές
- Παραδόσεις
- Εξυπηρέτηση

Αυτά τα αιτήματα των πελατών η εταιρία τα αντιμετωπίζει με ταχύτητα και αποτελεσματικότητα προσπαθώντας να μην γίνουν παράπονα.

Η υπηρεσία που προσφέρει η **Reevoo** σαν ανεξάρτητη αρχή καταγραφής, αξιολόγησης και προβολής των σχολίων των πελατών μας στο ηλεκτρονικό μας κατάστημα παίζει σημαντικό ρόλο. Τα σχόλια των πελατών χωρίζονται σε δύο κατηγορίες. Η πρώτη κατηγορία αφορά αξιολογήσεις που δέχεται το brand και η δεύτερη αφορά τα προϊόντα. Ο πελάτης που αγοράζει από τον Κωτσόβολο και μόνο έχει την δυνατότητα να τους αξιολογήσει στις δύο παραπάνω κατηγορίες. Η υπεύθυνοι της εταιρίας ανά τακτά χρονικά διαστήματα λαμβάνουν την σχετική αναφορά για το σύνολο των σχολίων που έχει δεχτεί η εταιρία σε κατηγορίες θετικά, αρνητικά, ουδέτερα.

Η Κωτσόβολος ελέγχει και παρακολουθεί τα σχόλια των χρηστών από τα Social Media και από το skroutz.gr. Είναι τα σημαντικότερα κανάλια επικοινωνίας που έχει ένας πελάτης με τους φίλους του οπότε είναι σημαντικό να μην υπάρχουν αρνητικά σχόλια πελατών σε αυτά.

6.2.3 Customer Plan



Εικόνα 20: Μηνιαία παρακολούθηση Customer Plan για την Κωτσόβολος

Με την εξυπηρέτηση των πελατών η Κωτσόβολος επιδιώκει να επιτύχει τα καλύτερα αποτελέσματα εξυπηρέτησης πελατών:

- Βελτιώνει την εξυπηρέτηση των πελατών συνεχώς
- Προστατεύει την εταιρία από την κρίση και εξοντώνει τον ανταγωνισμό
- Συνεχίζει τη βελτίωση των υποδομών
- Κερδίζει συνεχώς έδαφος από τα νέα κανάλια (Internet, B2B, Franchise)
- Καλύτερα καταστήματα σε εμπορικές θέσεις

7 Βιβλιογραφία

- [1] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.
- [2] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- [3] Liu, B. (2012). Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies). *Morgan & Claypool Publishers*.
- [4] Ogneva, M. (2010). How Companies Can Use Sentiment Analysis to Improve Their Business
- [5] Ganu, G., Elhadad, N., & Marian, A. (2009, June). Beyond the Stars: Improving Rating Predictions using Review Text Content. In *WebDB* (Vol. 9, pp. 1-6).
- [6] Pavlopoulos, J., & Androutsopoulos, I. (2014, April). Multi-granular aspect aggregation in aspect-based sentiment analysis. In *Proceedings of EACL* (pp. 78-87).
- [7] Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.
- [8] Pontiki, M., Papageorgiou, H., Galanis, D., Androutsopoulos, I., Pavlopoulos, J., & Manandhar, S. (2014, August). Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)* (pp. 27-35).
- [9] Suresh, V., Roohi, S., & Eirinaki, M. (2014, October). Aspect-based opinion mining and recommendationsystem for restaurant reviews. In *Proceedings of the 8th ACM Conference on Recommender systems* (pp. 361-362). ACM.
- [10] Carbonell, J. G. (1979). *Subjective Understanding: Computer Models of Belief Systems* (No. RR-150). YALE UNIV NEW HAVEN CONN DEPT OF COMPUTER SCIENCE.
- [11] Wilks, Y., & Bien, J. (1983). Beliefs, Points of View, and Multiple Environments*. *Cognitive Science*, 7(2), 95-119.
- [12] Hearst, M. A. (1992). Direction-based text interpretation as an information access refinement. *Text-based intelligent systems: current research and practice in information extraction and retrieval*, 257-274.

- [13] Huettner, A., & Subasic, P. (2000). Fuzzy typing for document management. *ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes*, 26-27.
- [14] Kantrowitz, M. (2003). *U.S. Patent No. 6,622,140*. Washington, DC: U.S. Patent and Trademark Office.
- [15] Sack, W. (1994, July). On the computation of point of view. In *AAAI* (p. 1488).
- [16] Wiebe, J. M., Bruce, R. F., & O'Hara, T. P. (1999, June). Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 246-253). Association for Computational Linguistics.
- [17] Wiebe, J. M., & Rapaport, W. J. (1988, June). A computational theory of perspective and reference in narrative. In *Proceedings of the 26th annual meeting on Association for Computational Linguistics* (pp. 131-138). Association for Computational Linguistics.
- [18] Cardie, C., Wiebe, J., Wilson, T., & Litman, D. J. (2003). Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. In *New directions in question answering* (pp. 20-27).
- [19] Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528). ACM.
- [20] Tong, R. M. (2001). An operational system for detecting and tracking opinions in on-line discussion. In *Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification* (Vol. 1, p. 6).
- [21] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.
- [22] Wiebe, J., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., ... & Maybury, M. T. (2003, March). Recognizing and Organizing Opinions Expressed in the World Press. In *New Directions in Question Answering* (pp. 12-19).
- [23] Yu, H., & Hatzivassiloglou, V. (2003, July). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing* (pp. 129-136). Association for Computational Linguistics.
- [24] Tatemura, J. (2000, January). Virtual reviewers for collaborative exploration of movie reviews. In *Proceedings of the 5th international conference on Intelligent user interfaces* (pp. 272-275). ACM.

- [25] Terveen, L., Hill, W., Amento, B., McDonald, D., & Creter, J. (1997). PHOAKS: A system for sharing recommendations. *Communications of the ACM*, 40(3), 59-62.
- [26] Stoyanov, V., Cardie, C., & Wiebe, J. (2005, October). Multi-perspective question answering using the OpQA corpus. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 923-930). Association for Computational Linguistics.
- [27] Cardie, C. (1997). Empirical methods in information extraction. *AI magazine*, 18(4), 65.
- [28] Meeker, M., & Wu, L. (2010). Internet trends. *Morgan Stanley*, 12.
- [29] Stock, J., Speh, T., & Shear, H. (2006). Managing product returns for competitive advantage. *MIT Sloan Management Review*, 48(1), 57-62.
- [30] Makriniwti, Vassalos. Supervised Sentiment Extraction from Greek tweets.