



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Πληροφορική»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Ανασκόπηση βιβλιογραφίας: Ανάλυση συναισθήματος σε εκπαιδευτικά συστήματα μέσω κοινωνικών δικτύων Literature review: Using Sentiment Analysis to enhance learning over social networks
Όνοματεπώνυμο Φοιτητή	ΔΜΟΥΛΗΣ ΝΙΚΟΛΑΟΣ
Πατρώνυμο	ΣΠΥΡΙΔΩΝ
Αριθμός Μητρώου	ΜΠΠΛ 14013
Επιβλέπων	MARIA VIRBOY

18 Νοεμβρίου 2016

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Όνομα Επώνυμο
Βαθμίδα

Όνομα Επώνυμο
Βαθμίδα

Όνομα Επώνυμο
Βαθμίδα

ΠΕΡΙΛΗΨΗ

Στην παρούσα εργασία γίνεται μια αναφορά στην ανάλυση συναισθημάτων στα κοινωνικά δίκτυα για εκπαιδευτικούς σκοπούς. Έτσι, παρουσιάζεται η χρήση και η εφαρμογή του web 2.0 καθώς και των social networks όπως επίσης και η εφαρμογή και χρήση της ανάλυσης συναισθήματος (sentiment analysis ή S.A) για εκπαιδευτικούς σκοπούς. Επιπλέον, περιγράφεται το υπόβαθρο και οι προσεγγίσεις με το S.A καθώς επίσης και μια ιδιαίτερη μελέτη σχετικά με το e- Learning. Τέλος, γίνεται μια αξιολόγηση και σύγκριση ορισμένων συστημάτων όπως τα blogs και το Twitter.

ABSTRACT

In this paper we refer the analysis of emotions in social networks for educational purposes. Thus, presented the use and application of web 2.0 and the social networks as well as the application and use of sentiment analysis (sentiment analysis or S.A) for educational purposes. Furthermore, described the background and approaches to S.A as well as a special study on e- Learning. Finally, there is an evaluation and comparison of some systems such as blogs and Twitter.

ΠΕΡΙΕΧΟΜΕΝΑ

Σύντομη Περιγραφή	4
Εισαγωγή.....	5
1. Προσεγγίσεις της ανάλυσης συναισθήματος	8
α) Υπόβαθρο.....	8
>Machine learning	8
>Lexicon-based method	10
>Hybrid approach	11
β) Αρχιτεκτονική ανάλυσης συναισθήματος.....	12
2. SA in e-learning	16
3. Comparative Analysis & Discussion	20
4. Conclusion & future work	25
Βιβλιογραφία.....	27

Σύντομη Περιγραφή

Τα τελευταία χρόνια οι νέες τεχνολογίες έχουν πλέον εφαρμογή σε κάθε τομέα της ζωής ενός σύγχρονου ανθρώπου. Ο χώρος της υγείας, ο επαγγελματικός και εκπαιδευτικός τομέας, οι επιστήμες, ακόμη και η καθημερινότητα και η επικοινωνία των ανθρώπων έχει πάρει άλλες διαστάσεις χάρη στις νέες τεχνολογίες. Είναι επομένως εύλογο το ενδιαφέρον που προκύπτει ώστε η παρούσα εργασία να ασχοληθεί με το παραπάνω ζήτημα. Πιο συγκεκριμένα, παρακάτω θα γίνει μια αναφορά στην ανάλυση συναισθημάτων στα κοινωνικά δίκτυα για εκπαιδευτικούς σκοπούς. Έτσι, παρουσιάζεται η χρήση και η εφαρμογή του web 2.0 καθώς και των social networks όπως επίσης και η εφαρμογή και χρήση της ανάλυσης συναισθήματος (sentiment analysis ή S.A) για εκπαιδευτικούς σκοπούς. Επιπλέον, περιγράφεται το υπόβαθρο και οι προσεγγίσεις με το S.A καθώς επίσης και μια ιδιαίτερη μελέτη σχετικά με το e- Learning. Τέλος, γίνεται μια αξιολόγηση και σύγκριση ορισμένων συστημάτων όπως τα blogs και το Twitter.



Εισαγωγή

Τα τελευταία χρόνια στο χώρο του διαδικτύου και της πληροφορικής έχει σημειωθεί μια σημαντική εξέλιξη και πλέον από το web 1.0 έχουμε περάσει στο web 2.0. Τα πλεονεκτήματα στις εφαρμογές του είναι ποικίλα και αφορούν πολλούς χρήστες, τόσο στον ιδιωτικό τομέα όσο και στον επαγγελματικό, σε οργανισμούς και εκπαιδευτικά ιδρύματα.

Αρχικά, η φράση web 2.0 ειπώθηκε για πρώτη φορά το 2004 στα πλαίσια ενός συνεδρίου όπου συμμετείχαν οι O'Reilly Media και MediaLive International προκειμένου να υπάρξει αναβάθμιση του διαδικτύου. Αφού παρατηρήθηκε ότι ο παγκόσμιος ιστός είχε αρχίσει να γίνεται πολύ δημοφιλής και να αποτελεί μέρος της καθημερινότητας των ανθρώπων, ενώ παράλληλα άρχισαν να παρουσιάζονται νέες εφαρμογές και ιστοσελίδες, οι οποίες σε πολύ σύντομο χρονικό διάστημα γίνονταν γνωστές από τους χρήστες, προέκυψε μια νέα γενιά υπηρεσιών μέσω διαδικτύου όπως ιστότοποι κοινωνικής δικτύωσης (social networks), wiki, εργαλεία επικοινωνίας και folksonomies με κύρια χαρακτηριστικά την ηλεκτρονική συνέργεια και την ανταλλαγή μεταξύ χρηστών. Έτσι οι δύο προαναφερόμενες εταιρείες εξέτασαν το web 2.0 ως τη νέα γενιά υπηρεσιών σχετικά με το διαδίκτυο και χρησιμοποίησαν αυτή τη φράση σε μια σειρά συνεδρίων τα οποία συνεχίζονται έως σήμερα προκειμένου να διασαφηνιστεί ο όρος. Παρόλα αυτά παρατηρείται πως από το 2005 ο συγκεκριμένος όρος χρησιμοποιείται ευρύτατα από τα άτομα που ασχολούνται με το διαδίκτυο. (wikipedia.org/wiki/Blog)

Έτσι λοιπόν, η χρήση του web 2.0 δεν έχει περιορισμούς και ο χρήστης έχει απεριόριστες δυνατότητες στον παγκόσμιο Ιστό. Πλέον παρέχεται η δυνατότητα να υπάρχει παρέμβαση του χρήστη στο περιβάλλον ιστοσελίδων και γενικότερα να γίνονται πολλές διαδράσεις. Χαρακτηριστικά παραδείγματα είναι το facebook και το Youtube όπου μπορεί κάποιος να εφαρμόσει διάφορες διαδράσεις όπως αναζήτηση (search), επεξεργασία συνδέσμων (linking), προσθήκη ετικέτας (tagging), επεξεργασία και διαγραφή πληροφοριών.

Κύρια χαρακτηριστικά του web 2.0 είναι τα εξής: (www.webz.gr/2007/11/13/web-20-goodies/)

- Υπάρχει ελεύθερη διακίνηση υπηρεσιών και δεδομένων από του χρήστες σε μια παγκόσμια πλατφόρμα η οποία αποτελείται από το διαδίκτυο και τις διάφορες συσκευές που συνδέονται με αυτό.
- Για να τρέξει μια web2.0 εφαρμογή χρειάζεται απλά ένας browser. Κάθε τέτοιου είδους εφαρμογή απαιτεί ύπαρξη του διαδικτύου και λειτουργεί μέσω διάφορων συσκευών όπως H/Y, κινητό τηλέφωνο, tablet.
- Περιέχει Λογισμικό, περιεχόμενο και εφαρμογές ανοιχτού κώδικα (open source).
- Χρήση κυρίως “ελαφριάς” τεχνολογίας σε ό,τι αφορά τα πρωτόκολλα, τις γλώσσες προγραμματισμού, τις διεπαφές χρήστη, ενώ διαπιστώνεται και μια τάση για απλότητα στον προγραμματιστικό σχεδιασμό τους.

- Πολυμεσικές και διαδραστικές διεπαφές χρήστη (Rich Internet Applications-RIA), δυναμικό περιεχόμενο, ιστοσελίδες που ανανεώνουν μόνο το περιεχόμενο που αλλάζει (τεχνολογία Ajax).
- Συνεχής και άμεση ανανέωση των δεδομένων και του λογισμικού που πρέπει να προσαρμόζεται διαρκώς στις ανάγκες των χρηστών.
- Προώθηση του δημοκρατικού χαρακτήρα του διαδικτύου, με τους χρήστες να έχουν τον πρωταγωνιστικό ρόλο.
- Υιοθέτηση της τάσης για αποκέντρωση των δεδομένων, υπηρεσιών και προτύπων.
- Δυνατότητα κατηγοριοποίησης του περιεχομένου από το χρήστη με σημασιολογικές έννοιες για ευκολότερη αναζήτηση της πληροφορίας.
- Δυνατότητα για ανοιχτή επικοινωνία, ανάδραση, διάχυση πληροφοριών, άμεση συγκέντρωση και εκμετάλλευση της γνώσης των χρηστών για διάφορα ζητήματα.
- Αμφίδρομη επικοινωνία του χρήστη με επιχειρήσεις ή οργανισμούς που μπορεί να έχει σαν αποτέλεσμα την υιοθέτηση κατευθύνσεων και τη λήψη αποφάσεων

Τα τελευταία χρόνια έχει παρατηρηθεί μια συνεχώς αυξανόμενη χρήση των Social Network. Με τον όρο αυτό ουσιαστικά γίνεται αναφορά σε κάθε site το οποίο δίνει τη δυνατότητα στους χρήστες μέσω πλατφόρμας να δημιουργήσουν κάποιο προφίλ και να αλληλεπιδράσουν με άλλους χρήστες μέσω «κοινωνικών συνδέσεων» εντός ενός χώρου ηλεκτρονικής κοινότητας. Χρησιμοποιώντας τον όρο “φιλία” οι χρήστες μπορούν να έχουν πρόσβαση ο ένας στο προφίλ του άλλου με τη δυνατότητα να μαθαίνουν για διάφορες ενημερώσεις και προσωπικά στοιχεία όπως π.χ φωτογραφίες. Εκτός από φίλους στα social networking sites υπάρχει και ο όρος "θαυμαστές" (fans-followers). Με την έννοια αυτή δηλώνεται η μονόδρομη σχέση επικοινωνίας μεταξύ των χρηστών του δικτύου. Θεωρητικά, οι δυνατότητες αλληλεπίδρασης είναι άπειρες και συνήθως περιορίζονται από τον χαρακτήρα που θέλει να εκφράσει το κάθε social networking site. Γενικά, τα social networks μπορούν να κατηγοριοποιηθούν σε δύο βασικές ομάδες, τα κάθετα social networks που περιλαμβάνουν χρήστες-μέλη με κοινά ενδιαφέροντα και κοινούς στόχους και τα οριζόντια social networks, που αποτελούνται από μέλη με διαφορετικά ενδιαφέροντα που συνήθως έχουν ως σκοπό απλά να έρθουν σε επικοινωνία μεταξύ τους, να γνωριστούν και να αλληλεπιδράσουν.(Martinez E. 2012). Ορισμένα παραδείγματα social networking site αποτελούν τα facebook, myspace και Twitter, τα οποία είναι ιδιαίτερα διαδεδομένα στους νέους και ειδικότερα το τελευταίο τίνει να χρησιμοποιείται λόγω ορισμένων χαρακτηριστικών που θα αναλυθούν παρακάτω, και στο χώρο της εκπαίδευσης.

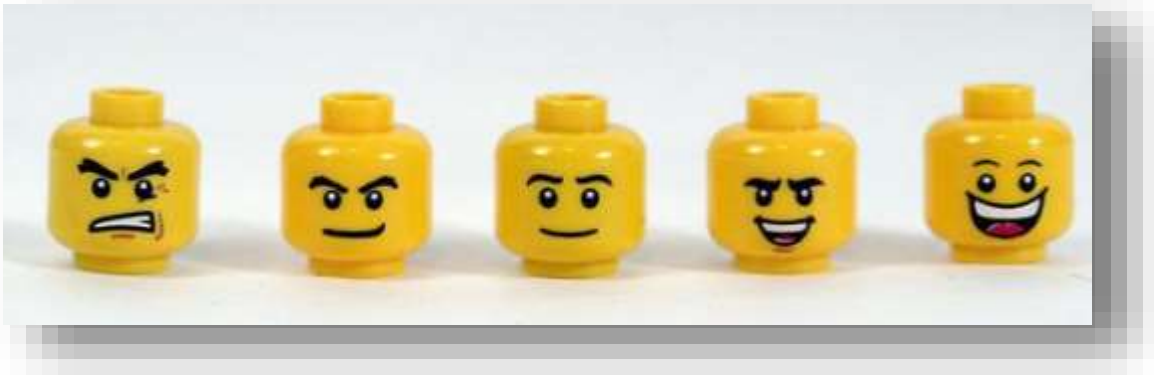
Ωστόσο, αξιόλογη είναι η προσέγγιση των ερευνητών που γίνεται για ανάλυση συναισθήματος στα social networking site. Σύμφωνα με το λεξικό της Οξφόρδης με την έννοια ανάλυση συναισθήματος(sentiment analysis) εννοείται: «Η διαδικασία της ταυτοποίησης και κατηγοριοποίησης των απόψεων που εκφράζονται σε ένα κείμενο, ειδικά

με σκοπό το να καθοριστεί εάν η άποψη του συγγραφέα σχετικά με ένα συγκεκριμένο θέμα, προϊόν κλπ είναι θετική, αρνητική ή ουδέτερη».(Schrauwen S., 2010), Ο Seth Grimes ένας από τους κορυφαίους industry analysts σε παγκόσμιο επίπεδο, δίνει τον εξής εναλλακτικό ορισμό: «Ως Ανάλυση Συναισθήματος ορίζουμε ένα σύνολο μεθόδων, οι οποίες τυπικά υλοποιούνται σε λογισμικό υπολογιστή, και ανιχνεύουν, μετράνε και αξιοποιούν στάσεις, συμπεριφορές, γνώμες και συναισθήματα σε online κοινωνικές και επιχειρηματικές πηγές πληροφορίας». (Cambria E. 2013)

Η Ανάλυση Συναισθήματος μπορεί να εφαρμοσθεί σε τρία επίπεδα: α) Σε επίπεδο εγγράφου (document-level SA), όπου υποθέτει ότι κάθε κείμενο εκφράζει μία και μοναδική άποψη για ένα συγκεκριμένο θέμα ή αντικείμενο. Και περιορίζεται σε θετική ή αρνητική κλάση β) Σε επίπεδο πρότασης (sentence-level SA), όπου κάθε έγγραφο χωρίζεται σε προτάσεις, υποθέτοντας ότι καθεμιά από αυτές εκφράζει μια ξεχωριστή άποψη και περιορίζεται σε θετική, αρνητική και ουδέτερη κλάση και γ) Σε επίπεδο χαρακτηριστικών (feature-level SA), όπου κάθε έγγραφο χωρίζεται σε προτάσεις και κάθε πρόταση σε πολωμένες φράσεις που αντιστοιχούν σε ένα συγκεκριμένο χαρακτηριστικό του πραγματευόμενου θέματος ή αντικειμένου. (Mr S. Vohra et al.2016)

Με βάση λοιπόν τα παραπάνω, η παρούσα εργασία αρχικά στο πρώτο κεφάλαιο παρουσιάζει τις διάφορες προσεγγίσεις του S.A και έπειτα γίνεται αξιολόγηση των προσεγγίσεων που παρουσιάζονται σε συγκεκριμένα άρθρα. Ακολούθως, στο δεύτερο κεφάλαιο, γίνεται μια αναφορά στη χρήση του S.A στο e-learning, ενώ στο τρίτο κεφάλαιο που ακολουθεί, γίνεται σύγκριση των συστημάτων του S.A που έχουν αναφερθεί νωρίτερα βάση ορισμένων κριτηρίων. Τέλος, παρουσιάζονται τα συμπεράσματα που προκύπτουν από τη συγκεκριμένη μελέτη.

1. Προσεγγίσεις της ανάλυσης συναισθήματος



α) Υπόβαθρο

Οι πρώτες προσπάθειες στην ανάλυση συναισθήματος παρουσιάστηκαν τη δεκαετία του 70 αλλά ουσιαστικά απογειώθηκαν με τη χρήση του Web 2.0. Οι πιο συνηθισμένες προσεγγίσεις της ανάλυσης συναισθήματος είναι η Μηχανική μάθηση, (machine learning), το λεξικό (lexicon-based method) και οι υβριδικές προσεγγίσεις (hybrid approach).

>Machine learning

Η Μηχανική Μάθηση είναι μια κατηγορία έξυπνης θα λέγαμε τεχνολογίας που ασχολείται με αλγόριθμους που επιτρέπουν στους υπολογιστές να μαθαίνουν. Σύμφωνα με έναν πιο τυπικό ορισμό είναι η αυτοματοποιημένη διαδικασία εξαγωγής προτύπων ή μοτίβων (patterns) από μεγάλους όγκους δεδομένων και χρησιμοποίησής των για τη διεξαγωγή προβλέψεων πάνω σε νέα δεδομένα. Συνήθως όταν υλοποιούμε συστήματα μηχανικής μάθησης χρησιμοποιούμε ένα σύνολο δεδομένων εκπαίδευσης. Το σύνολο αυτό είναι τα δεδομένα τα οποία χρησιμοποιεί ο αλγόριθμος για να κατανοήσει τη δομή των δεδομένων και να βρει το υποβόσκον μοντέλο που θα χρησιμοποιηθεί αργότερα για το χειρισμό νέων, άγνωστων δεδομένων. Η ικανότητα του σωστού χειρισμού. (Medhat W. et al, 2014)

Η ικανότητα του σωστού χειρισμού άγνωστων δεδομένων βάση του συνόλου εκπαίδευσης, ονομάζεται **ικανότητα γενίκευσης (generalization)**. Η υπερπροσαρμογή (overfitting) από την άλλη μεριά, συμβαίνει όταν ένα σύνολο δεδομένων εκπαίδευσης έχει οδηγήσει σε ένα μοντέλο το οποίο είναι υπερβολικά προσαρμοσμένο πάνω στο συγκεκριμένο σύνολο δεδομένων, καθιστώντας αδύνατη τη γενίκευσή του. Η υπερ-εκπαίδευση του συνόλου δεδομένων μπορεί να αποφευχθεί αξιολογώντας το μοντέλο με κάποια δεδομένα (test data), τα οποία μπορούν να είναι ένα υποσύνολο των διαθέσιμων δεδομένων. Όταν έχουμε να επιλέξουμε μεταξύ πολλών ταξινομητών, αυτός που αποδίδει καλύτερα πάνω στο testing set έχει την μεγαλύτερη πιθανότητα να είναι αυτός με την καλύτερη ικανότητα γενίκευσης (Saif H, et al, 2012). Τα μοντέλα μηχανικής μάθησης συνήθως χωρίζονται σε τρεις μεγάλες κατηγορίες ανάλογα με την επίβλεψη που απαιτείται κατά τη διάρκεια της εκπαίδευσης:

Στην επιβλεπόμενη μάθηση (supervised learning) η εκπαίδευση των ταξινομητών γίνεται με επισημειωμένα δείγματα (labelled samples), δηλαδή, στην περίπτωση του Twitter, με tweets που έχουν επισημειωθεί με την κλάση που ανήκουν (αρνητική, θετική ή ουδέτερη). Οι Naïve Bayes, Maximum Entropy και Support Vector Machines (SVM) είναι οι πιο γνωστοί ταξινομητές αυτής της κατηγορίας (Kotsiantis S. et al ,2007)

Στη μη-επιβλεπόμενη μάθηση (unsupervised learning) δουλεύουμε με μη- επισημειωμένα δεδομένα (unlabeled data). Διάσημοι μη-επιβλεπόμενοι αλγόριθμοι περιλαμβάνουν τους αλγόριθμους ομαδοποίησης (clustering algorithms) όπως k-means, k-methods, hierarchical clustering, τα κρυφά μοντέλα Markov (HMM Hidden Markov Models) και κάποια μη επιβλεπόμενα μοντέλα νευρωνικών δικτύων, όπως οι αυτοοργανούμενοι χάρτες.(Medhat W. et al, 2014)

Στην ημι-επιβλεπόμενη μάθηση (semi-supervised learning) η εκπαίδευση των ταξινομητών γίνεται τόσο με επισημειωμένα όσο και με μη- επισημειωμένα δεδομένα. Τέτοιοι αλγόριθμοι είναι ο label propagation και τα διάφορα μοντέλα βασισμένα σε γράφους (graph-based models).

Αναλύοντας μια μεγάλη ποσότητα κειμένων, ένα μοντέλο μπορεί να εκπαιδευτεί να ταξινομεί νέα κείμενα με βάση τις ομοιότητες με τα κείμενα με τα οποία εκπαιδεύτηκε. Η ταξινόμηση απαιτεί τα κείμενα εκπαίδευσης να είναι ήδη επισημειωμένα (annotated) με κάποιο συναίσθημα, έτσι ώστε ο αλγόριθμος να προσπαθήσει να βρει τι τα διαχωρίζει και να χρησιμοποιήσει αυτή τη γνώση σε νέα, άγνωστα κείμενα. Το πλεονέκτημα των μεθόδων μηχανικής μάθησης είναι ότι δεν απαιτείται κάποια γνώση σχετικά με τη γλώσσα στην οποία είναι γραμμένα τα κείμενα. Ιδανικά, κάθε τέτοιο σύστημα θα μπορούσε να χρησιμοποιηθεί με πολλαπλές γλώσσες, κάτι που είναι πολύ χρήσιμο για Ανάλυση Συναισθήματος σε κείμενα γραμμένα σε γλώσσες για τις οποίες δεν υπάρχουν διαθέσιμα αξιόπιστα εργαλεία επεξεργασίας φυσικής γλώσσας (NLP tools). (Medhat W. et al, 2014)

Αν και οι επιβλεπόμενες τεχνικές μηχανικής μάθησης έχουν αποδειχτεί οι πιο χρήσιμες και αποτελεσματικές για την Ανάλυση Συναισθήματος, η ιδιαίτερη φύση του Twitter και των μηνυμάτων του θέτουν αρκετούς περιορισμούς σε αυτού του είδους την προσέγγιση. Καταρχάς, εξαρτώνται από την ύπαρξη εκτεταμένων επισημειωμένων συλλογών από tweets για την εκπαίδευση των ταξινομητών, την ίδια στιγμή που η διαδικασία επισημείωσης κειμένων είναι συνήθως αρκετά ακριβή και χρονοβόρα , ειδικά για συνεχώς μεταβαλλόμενα και εξελισσόμενα θεματικά πεδία όπως στο Twitter. Δεύτερον, οι επιβλεπόμενες τεχνικές είναι συνήθως εξαρτώμενες από το εκάστοτε πεδίο (domain dependent), δηλαδή οι ταξινομητές που εκπαιδεύονται σε δεδομένα από ένα συγκεκριμένο πεδίο μπορεί να εμφανίσουν χαμηλή απόδοση όταν εφαρμοστούν σε δεδομένα από άλλο θεματικό πεδίο . Τρίτον τα tweets τείνουν να είναι πολύ αραιά εξαιτίας της συχνής χρήσης συντομογραφιών, ανορθογραφιών, γραμματικά λανθασμένων λέξεων και μη τυπικών εκφράσεων, γεγονός που επηρεάζει αρνητικά την απόδοση του ταξινομητή καθώς πολλοί όροι των δεδομένων εκπαίδευσης δεν εμφανίζονται στα δεδομένα αξιολόγησης. Οι περισσότερες προσπάθειες για να

ξεπεραστούν οι παραπάνω δυσκολίες έχουν να κάνουν με τη χρήση μη-επιβλεπόμενων ταξινομητών για τη μείωση της εξάρτησης από επισημειωμένα δεδομένα με τη χρήση διαφορετικών διαδικασιών (feature engineering processes) και την υιοθέτηση τεχνικών μείωσης διάστασης (dimensionality reduction techniques) για τη μείωση της αραιότητας των tweets (Kolchyna O. et al, 2015).

>Lexicon-based method

Η προσέγγιση βασισμένη σε λεξικό (lexicon-based method) υποθέτει ότι ο συναισθηματικός προσανατολισμός ενός κειμένου μπορεί να συναχθεί από το συναισθηματικό προσανατολισμό των επιμέρους λέξεων και φράσεων του. Σε αντίθεση με τις μεθόδους βασισμένες σε μηχανική μάθηση, η προσέγγιση βάση λεξικού δεν απαιτεί την εκπαίδευση ενός ταξινομητή. Αντιθέτως, χρησιμοποιεί λεξικά συναισθήματος για να αποδώσει το συναίσθημα των συναισθηματικά φορτισμένων λέξεων στο κείμενο. Η απόδοση μιας μεθόδου Ανάλυσης Συναισθήματος βασισμένη σε λεξικό συνήθως καθορίζεται από τον τύπο του λεξικού συναισθήματος και από τον αλγόριθμο ανίχνευσης συναισθήματος (sentiment detection algorithms), δηλαδή από τον αλγόριθμο που χρησιμοποιείται για τον εντοπισμό των συναισθηματικά φορτισμένων λέξεων του κειμένου και τον υπολογισμό του συνολικού συναισθήματος. (Mr M. Vohra et al., 2016, Medhat W. et al, 2014) Σε ό,τι αφορά το Twitter η πιο συνηθισμένη πρακτική είναι η χρήση ενός προενσωματωμένου λεξικού συναισθήματος μαζί με έναν απλό αλγόριθμο ταυρίσματος λέξεων κλειδίων (keywords). (Kolchyna O. et al, 2015)

Οι βασισμένες σε λεξικό μέθοδοι έχουν το πλεονέκτημα της απλότητας και της ταχύτητας, στοιχεία που αποτελούν απαραίτητη προϋπόθεση ιδίως όταν απαιτείται η ανάλυση τεράστιων όγκων δεδομένων, όπως για παράδειγμα σε συλλογές που αποτελούνται από δεκάδες χιλιάδες tweets. Από την άλλη, οι μέθοδοι αυτοί έρχονται συχνά αντιμέτωπες με δύο βασικούς περιορισμούς:

1) Οι συμβατικές μέθοδοι αδυνατούν να ανιχνεύσουν σύνθετους τύπους συναισθήματος σε tweets, όπως για παράδειγμα την άρνηση ή το ενισχυμένο συναίσθημα. Επιπλέον εξαιτίας του γεγονότος ότι στα λεξικά συναισθήματος οι λέξεις επισημειώνονται με μια σταθερή πολικότητα, δεν λαμβάνεται υπόψη το ευρύτερο σημασιολογικό πλαίσιο (context).

2) Τα παραδοσιακά λεξικά, όπως τα MPQA και SentiWordNet για την αγγλική γλώσσα, είναι σχεδιασμένα για να χειρίζονται επίσημα και καλογραμμένα κείμενα. Στο Twitter όμως, ο ιδιαίτερος και ανεπίσημος τρόπος γραφής εξαιτίας του περιορισμού των 140 χαρακτήρων, οδηγεί στον εντοπισμό ανάκτηση ελάχιστων μη-συμβατικών ή ασυνήθιστων λέξεων, αφού αυτές σπάνια είναι καταχωρημένες στα παραδοσιακά λεξικά.

Κατά συνέπεια, οι μέθοδοι βασισμένες σε λεξικό μπορούν να εφαρμοστούν άμεσα σε δεδομένα από το Twitter, χωρίς να χάνεται πολύτιμος χρόνος για την εκπαίδευση ταξινομητών, με το βασικό περιορισμό ότι η απόδοση και αποτελεσματικότητα όλου του συστήματος συνδέεται αναπόσπαστα με τη λεξική πηγή στην οποία βασίζεται. Μια μέθοδος βασισμένη σε λεξικό είναι τόσο καλή, όσο και το λεξικό που χρησιμοποιεί. Τέλος, η απόδοση τέτοιων συστημάτων, σε επίπεδο ακρίβειας και χρονικής πολυπλοκότητας, επιδεινώνεται δραστικά με την εκθετική αύξηση του μεγέθους του λεξικού. (Mr M. Vohra et al. , 2016)

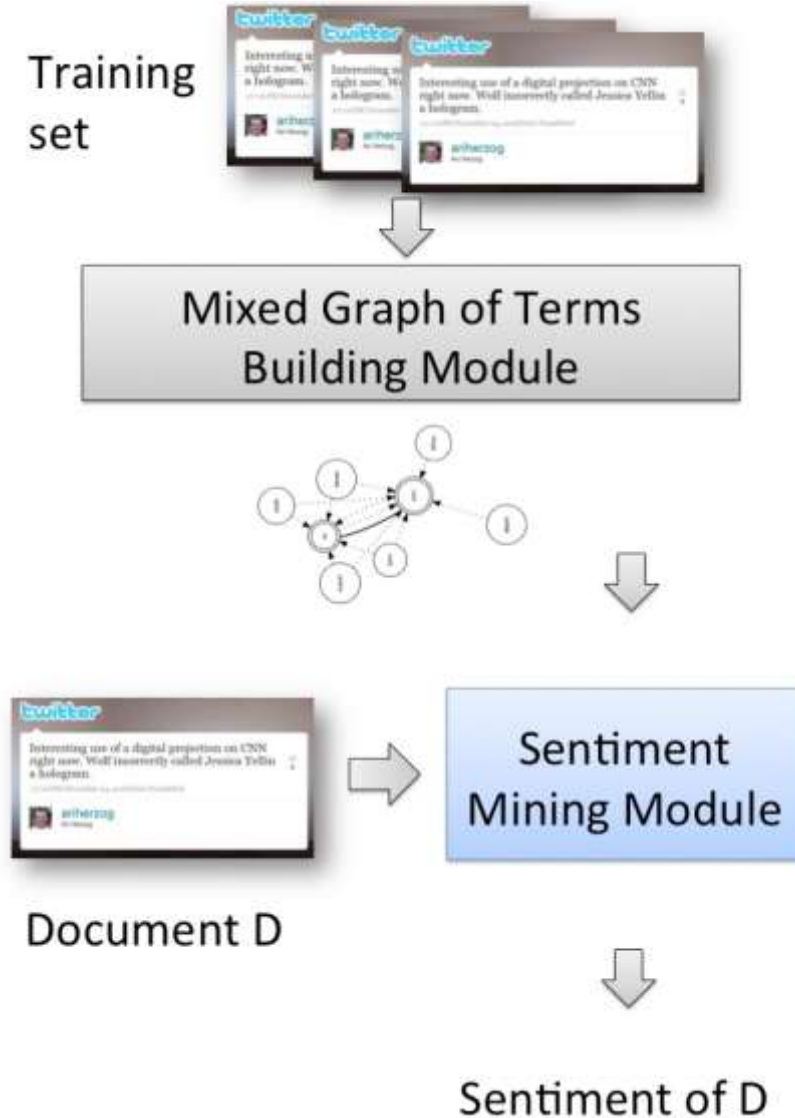
Στο σημείο αυτό αξίζει να σημειωθεί ότι σε οποιαδήποτε εφαρμογή ανάλυσης κειμένου υπάρχει η δυνατότητα επιλογής μεταξύ στατιστικών ή συντακτικών τεχνικών. Οι συντακτικές τεχνικές (syntactic techniques) μπορούν να οδηγήσουν σε καλύτερη ακρίβεια γιατί κάνουν χρήση των συντακτικών κανόνων μιας γλώσσας με σκοπό να ανιχνεύσουν τα ρήματα, τα επίθετα και τα ουσιαστικά. Δυστυχώς αυτού του είδους οι τεχνικές είναι άρρηκτα συνδεδεμένες με τη γλώσσα του κειμένου και συνεπώς δεν μπορούν να εφαρμοστούν σε άλλες γλώσσες. Από την άλλη οι στατιστικές τεχνικές (statistical techniques) έχουν πιθανό υπόβαθρο και εστιάζουν στις σχέσεις μεταξύ των λέξεων και

των κατηγοριών. Οι τεχνικές αυτές έχουν δύο σημαντικά πλεονεκτήματα σε σχέση με τις συντακτικές: μπορούμε να τις χρησιμοποιήσουμε σε διάφορες γλώσσες τροποποιώντας τις ελάχιστα ή και καθόλου και μπορούμε να χρησιμοποιήσουμε μετάφραση μηχανής (machine translation) του αρχικού σετ δεδομένων και να έχουμε και πάλι ικανοποιητικά αποτελέσματα. Οι δύο προαναφερθείσες μέθοδοι μπορούν να συνδυαστούν τόσο με τεχνικές βασισμένες σε λεξικό, όσο και με τεχνικές μηχανικής μάθησης. (Medhat W. et al, 2014)

>Hybrid approach

Η συνεχώς αυξανόμενη πρόοδος που παρατηρείται στον τομέα της Ανάλυσης Συναισθήματος έχει οδηγήσει τους ερευνητές στη μελέτη πιθανότητας μιας υβριδικής προσέγγισης η οποία συνολικά θα παρουσίαζε την ακρίβεια των μεθόδων μηχανικής μάθησης και την ταχύτητα των μεθόδων βασισμένων σε λεξικά συναισθήματος. Στην ουσία αυτό που προσπαθούν οι υβριδικές προσεγγίσεις είναι να εκμεταλλευτούν τα πλεονεκτήματα των δύο μεθόδων και να αποφύγουν τα μειονεκτήματα. Στη γενική περίπτωση, στις υβριδικές προσεγγίσεις, η μία από τις δύο βασικές προσεγγίσεις χρησιμοποιείται για να τονώσει την απόδοση της άλλης προσέγγισης. Για παράδειγμα, από τη μία μεριά οι επιβλεπόμενοι ταξινομητές μπορεί να κάνουν χρήση μεθόδων βασισμένων σε λεξικό για να μειώσουν την εξάρτηση από χειροκίνητες επισημειωμένες συλλογές δεδομένων εκπαίδευσης, ενώ από την άλλη οι αλγόριθμοι μηχανικής μάθησης μπορούν να χρησιμοποιηθούν για να «στηρίξουν» (bootstrap) τα λεξικά συναισθήματος στις μεθόδους βασισμένες σε λεξικό. (Mr M. Vohra et al. , 2016)

β) Αρχιτεκτονική ανάλυσης συναισθήματος



Εικόνα

Είναι γεγονός ότι σήμερα μέσα από τα Μέσα Κοινωνικής Δικτύωσης μπορεί κανείς να αντλήσει πολλές πληροφορίες. Μια βασική ανάγκη που δημιουργείται είναι να γίνει εκμετάλλευση αυτής της μεγάλης ποσότητας δεδομένων και να ανακαλυφθεί η γνώση που κρύβεται μέσα σε αυτά. Η Εξόρυξη Γνώσης ή Ανακάλυψη Γνώσης από Βάσεις Δεδομένων ή Εξόρυξη Δεδομένων (Data Mining - DM) (Martinez et al, 2012) είναι η εξεύρεση μιας ενδιαφέρουσας, αυτονόητης, μη προφανούς και πιθανόν χρήσιμης πληροφορίας ή προτύπων από μεγάλες βάσεις δεδομένων με χρήση αλγορίθμων ομαδοποίησης ή κατηγοριοποίησης και των αρχών της στατιστικής, της

τεχνητής νοημοσύνης, της μηχανικής μάθησης και των συστημάτων βάσεων δεδομένων. Στόχος της εξόρυξης δεδομένων είναι η πληροφορία που θα εξαχθεί και τα πρότυπα που θα προκύψουν να έχουν δομή κατανοητή προς τον άνθρωπο έτσι ώστε να τον βοηθήσουν να πάρει τις κατάλληλες αποφάσεις.

Μια υποκατηγορία της εξόρυξης γνώσης είναι η εξόρυξη κειμένου (Text Mining - TM), η οποία ορίζεται σαν μια διαδικασία εξαγωγής με αυτόματο τρόπο νέας, έγκυρης, και χρήσιμης γνώσης από διαφορετικούς γραπτούς πόρους, καθώς επίσης και η όσο το δυνατόν καλύτερη οργάνωση αυτής της νέας γνώσης-πληροφορίας για την όποια μελλοντική αναφορά (Martinez et al, 2012; Sehgal, 2004). Μια υποκατηγορία της εξόρυξης κειμένου είναι η εξόρυξη γνώμης (opinion mining -OM) ή ανάλυση συναισθήματος (sentimental analysis) η οποία στοχεύει να εξαγάγει γνωρίσματα και συστατικά ενός αντικειμένου που έχει σχολιαστεί και να καθορίσει αν το σχόλιο είναι θετικό, αρνητικό ή ουδέτερο (Liu, 2012).

Εφαρμόζοντας τις τεχνικές εξόρυξης γνώσης στα Κοινωνικά Δίκτυα είναι σε θέση κάποιος να ανακαλύψει ποικίλες πλευρές της ανθρώπινης συμπεριφοράς και της ανθρώπινης αλληλεπίδρασης, να βελτιωθεί η αντίληψη που έχουν οι άνθρωποι σχετικά με ένα θέμα, να προσδιοριστούν ομάδες ανθρώπων ανάμεσα στις μάζες του πληθυσμού, να μελετηθούν ομάδες που αλλάζουν με το χρόνο, να βρεθούν άνθρωποι με επιρροή, ή ακόμα και να γίνει η σύσταση ενός προϊόντος ή μιας δραστηριότητας σε ένα άτομο. (Ταουφίq, 2016). Επειδή τα ποσά δεδομένων είναι τεράστια, είναι απαραίτητο να εφαρμοστεί μια διαδικασία αυτόματης κατηγοριοποίησης των κειμένων. Αρχικά συλλέγεται ένα σύνολο κειμένων, το οποίο κάποιος ειδικός ή μια ομάδα ανθρώπων έχουν σημασιοδοτήσει (labeled) σε σχέση με κάποιες κατηγορίες. Το σύνολο των κειμένων χωρίζεται σε ένα σύνολο εκπαίδευσης (training set), και ένα σύνολο ελέγχου (test set) το οποίο χρησιμοποιείται για επικύρωση (validation) της απόδοσης του αλγορίθμου κατηγοριοποίησης.

Στο σύνολο εκπαίδευσης, το συνολικό κείμενο μετατρέπεται σε ξεχωριστές λέξεις. Όλες οι λέξεις που περιλαμβάνονται στα κείμενα μπαίνουν σε «σακούλα λέξεων» (bag-of-words). Παράλληλα, αγνοούνται όλοι οι μη αφηριθμητικοί όροι, π.χ. αριθμοί, ημερομηνίες, σύμβολα όπως “, ‘=’ κ.α. Στη συνέχεια οι λέξεις μετατρέπονται στη ρίζα τους (stemming). Έπειτα αφαιρούνται οι κοινές λέξεις (άρθρα, επιρρήματα, κτλ). Έχει παρατηρηθεί πως οι 10 συχνότερες λέξεις της αγγλικής γλώσσας αποτελούν το 20-30% των λεκτικών μονάδων σε ένα κείμενο (Chien-wen Shen, 2015). Ανάμεσά τους οι λέξεις {is, the, to, for, and, it...} που συναντάμε σχεδόν σε κάθε πρόταση. Η ποσότητα των όρων που αποφασίζουμε πως δε μεταφέρουν πληροφορία επηρεάζει και την ποιότητα των δεδομένων που θα αναπαραστήσουμε. (Keethi et al, 2014)

Έπειτα γίνεται διανυσματική αναπαράσταση του κειμένου (vector representation). Έτσι, κάθε αρχείο κειμένου από το σύνολο κειμένων που έχουμε είναι και ένα διάνυσμα όρων (term vector) στο οποίο κάθε όρος αποτελεί ένα μοναδικό ανεξάρτητο χαρακτηριστικό (feature). Κάθε στοιχείο σε αυτό το διάνυσμα έχει και μια τιμή η οποία αντιστοιχεί στην εμφάνιση του όρου μέσα στο κείμενο. Μια συχνά χρησιμοποιούμενη μέθοδος που δίνει πολύ καλά αποτελέσματα είναι η συνάρτηση TF-IDF (Term Frequency Inverse Document Frequency) (Sehgal, 2004). Τέλος χρησιμοποιώντας το σύνολο εκπαίδευσης, ένας επιβλεπόμενος αλγόριθμος εκμάθησης (supervised learning algorithm) προσπαθεί να διαμορφώσει το βέλτιστο ταξινομητή. (Keethi et al, 2014)

Συγκεκριμένα, κάνοντας λόγο για την ανάλυση συναισθήματος, όπως αναφέρθηκε στην προηγούμενη ενότητα, προσεγγίζεται με δύο διαφορετικούς τρόπους. Ο πρώτος αναφέρεται ως lexicon-based και κάνει χρήση συναισθηματικών λεξικών για να αποδώσει συναισθηματική βαθμολογία σε λέξεις και φράσεις. Στη συνέχεια συνθέτει τις λέξεις και τις φράσεις για να προκύψει συναισθηματική βαθμολογία για όλο το προς εξέταση κείμενο. Μία απλή υλοποίηση της lexicon-based προσέγγισης, στην περίπτωση απλών προτάσεων, είναι η πρόσθεση των συναισθηματικών βαθμολογιών των λέξεων με τον τρόπο που υποδεικνύει το συντακτικό δέντρο της πρότασης. Έτσι

Ξεκινώντας από τα φύλλα, που αποτελούν λέξεις, γίνονται διαδοχικές αθροίσεις των βαθμολογιών μέχρι το τελικό σκορ στη ρίζα του δέντρου, που χαρακτηρίζει όλη την πρόταση. Ανάλογα με την τελική βαθμολογία η πρόταση χαρακτηρίζεται αρνητική ή θετική. Τέτοιες προσεγγίσεις συχνά περιέχουν και χειρισμό της άρνησης (negation handling). Καθώς η άρνηση αντιστρέφει το συναίσθημα της λέξης (cool - not cool), μπορεί για παράδειγμα η συναισθηματική βαθμολογία λέξεων που συντάσσονται με άρνηση απλά να αλλάζει πρόσημο. Διατηρείται δηλαδή η ένταση του συναισθήματος αλλά αντιστρέφεται η πολικότητα..

Ο δεύτερος τρόπος προσέγγισης του προβλήματος της ανάλυσης συναισθήματος είναι η εξαγωγή χαρακτηριστικών και η χρήση μεθόδων μηχανικής μάθησης. Όπως αναφέρθηκε προηγουμένως, η μηχανική μάθηση είναι ο κλάδος της επιστήμης των υπολογιστών που μελετά και κατασκευάζει αλγόριθμους και τεχνικές που δίνουν τη δυνατότητα στον υπολογιστή να μαθαίνει από την εμπειρία. Βασικό χαρακτηριστικό της μηχανικής μάθησης είναι ότι αναζητεί πρότυπα (patterns) και σχέσεις στα δεδομένα με σκοπό να τα μοντελοποιήσει και να κάνει προβλέψεις πάνω σε αυτά.

Ένα από τα εργαλεία προσέγγισης που χρησιμοποιείται στη μηχανική μάθηση είναι η θεωρία απόφασης του Bayes, η οποία αποτελεί μια από τις σημαντικότερες στατιστικές προσεγγίσεις για το πρόβλημα της ταξινόμησης προτύπων. Βασίζεται στη σύγκριση μεταξύ διαφόρων αποφάσεων ταξινόμησης με βάση τις πιθανότητες και τα κόστη που σχετίζονται με τις αποφάσεις αυτές. Θεωρεί ότι το πρόβλημα απόφασης ορίζεται με πιθανό θεωρητικούς όρους και ότι όλες οι σχετικές πιθανότητες είναι γνωστές. Στην βασική του μορφή το θεώρημά του Bayes είναι απλό και αφορά υπό συνθήκη πιθανότητες: Αν A και B είναι δύο ενδεχόμενα με $P(A) > 0$, τότε: $P(B | A) = P(A | B)P(B) / P(A)$

Ο απλοϊκός ταξινομητής Bayes, παρά τις παραδοχές ανεξαρτησίας είναι ένας εξαιρετικά δημοφιλής ταξινομητής και σημειώνει καλές επιδόσεις σε πολλά προβλήματα του πραγματικού κόσμου. Επίσης εκπαιδεύεται αποδοτικά εξαιτίας της απλότητάς του. Η εκπαίδευσή του, όπως προαναφέρθηκε, συνιστά ουσιαστικά τη στατιστική εκτίμηση των συναρτήσεων πυκνότητας πιθανότητας και μετά την θεώρηση της ανεξαρτησίας των χαρακτηριστικών. Η προσέγγιση αυτή βασίζεται στη θεώρηση κάποιας τυπικής κατανομής για τα δεδομένα και την εκτίμηση των παραμέτρων της κατανομής με τη μέθοδο μέγιστης πιθανοφάνειας (maximum likelihood - ML). (Martinez et al, 2012)

Από την άλλη πλευρά, Η μηχανή διανυσμάτων υποστήριξης (support vector machine - SVM) είναι ένας γραμμικός ταξινομητής που διαχωρίζει δεδομένα δύο κλάσεων σε ένα χώρο πολλών διαστάσεων. Χρησιμοποιείται σε προβλήματα ταξινόμησης και στην προσέγγιση της μορφής της συνάρτησης σε προβλήματα παλινδρόμησης. Η λογική μια μηχανής εκμάθησης είναι να δίνει την τιμή y μιας συνάρτησης (άγνωστη προς εμάς) που αντιστοιχεί σε δοσμένο σημείο x. Αυτό γίνεται ως εξής: Για δεδομένο σύνολο σημείων x και έχοντας τις αντίστοιχες τιμές y που παίρνει η άγνωστη συνάρτηση, εκπαιδεύουμε τη μηχανή εκμάθησης να μάθει τη σχέση που συνδέει τα x με τα y. Δηλαδή, η μηχανή μαθαίνει την αντιστοίχιση $x \rightarrow y$ και έτσι για ένα σημείο x, διαφορετικό από αυτά του συνόλου της εκμάθησης, θα μας δώσει την τιμή y που θα έπαιρνε η άγνωστη συνάρτηση. Στην περίπτωση ταξινόμησης με τα SVM, το σύνολο των σημείων l αποτελείται από δύο υποσύνολα τα k και n. Έτσι, το αποτέλεσμα της συνάρτησης θα είναι +1 ή -1 ($y_i = +1$ ή $y_i = -1$) ανάλογα σε ποιο υποσύνολο ανήκει το δοθέν σημείο x. Τα δύο αυτά υποσύνολα ονομάζονται κλάσεις και η τιμή +1 (-1) είναι η "ετικέτα" της κλάσης. Δηλαδή, σε αυτή τη περίπτωση τα SVM μαθαίνουν να κατατάσσουν σωστά τα σημεία i x r στις δύο κλάσεις. Τα σημεία i x r και οι αντίστοιχες τιμές τους, y, αποτελούν την πληροφορία εκπαίδευσης (training set). Τα σημεία x ονομάζονται πρότυπα εκπαίδευσης (training patterns) ενώ οι τιμές y που αντιστοιχούν σε αυτά, στόχοι εκπαίδευσης (training targets). (Chen et al, 2013)

Ωστόσο, ένα πρόβλημα των εφαρμογών επιβλεπόμενης μηχανικής μάθησης, είναι η εύρεση καλών, ποιοτικά και ποσοτικά, συνόλων δεδομένων που συνοδεύονται από επιθυμητές αποκρίσεις. Τέτοια σύνολα δεδομένων καλούνται *labeled datasets* και είναι απαραίτητα στην επιβλεπόμενη μάθηση εν αντιθέσει με την μη επιβλεπόμενη μάθηση που χρησιμοποιεί *unlabeled* δεδομένα. Η προσάρτηση επιθυμητών εξόδων σε δεδομένα απαιτεί στις περισσότερες των περιπτώσεων ανθρώπινη εργασία, δηλαδή ανθρώπους να χαρακτηρίζουν τα δεδομένα ένα προς ένα ως προς το περιεχόμενό τους. Αυτό, όπως είναι φυσικό, θέτει περιορισμούς τόσο στην ποιότητα όσο και στην ποσότητα των δεδομένων. Εξαιτίας του μη αυτοματοποιημένου τρόπου εξαγωγής των *labels* (επιθυμητές έξοδοι), *labeled datasets* με μεγάλο αριθμό δεδομένων είναι δύσεύρετα. Επίσης, πολλές φορές το περιεχόμενο των δεδομένων δεν είναι άμεσα προφανές ούτε στον άνθρωπο, με αποτέλεσμα να υπάρχουν ασυμφωνίες στο χαρακτηρισμό των δεδομένων (*labeling of data*) και να απαιτούνται περισσότεροι του ενός άνθρωποι (*annotators*) για αυτό τον χαρακτηρισμό.

Η υπερπληθώρα δεδομένων στο διαδίκτυο, συνιστά κυρίως *unlabeled* δεδομένα, ωστόσο το *web2.0* δίνει δυνατότητες για την πλήρη ή μερική αυτοματοποίηση της διαδικασίας χαρακτηρισμού δεδομένων σε ορισμένες περιπτώσεις. Στο πεδίο της ανάλυσης συναισθήματος, για παράδειγμα, όταν εξετάζεται η συναισθηματική πολικότητα σε κριτικές από διαδικτυακές πηγές, η αυτοματοποίηση της διαδικασίας χαρακτηρισμού είναι εύκολη. Καθώς υπάρχουν πολλές διαδικτυακές πηγές, όπου χρήστες ανεβάζουν κριτικές προϊόντων και συνοδεύουν τις κριτικές αυτές από κάποια βαθμολογία (*star rating*) ο χαρακτηρισμός δεδομένων, επιμερίζεται σε όλους τους χρήστες και όχι σε μερικούς *annotators*. Έτσι στο διαδίκτυο συναντά κανείς πολλά *datasets* με κριτικές ταινιών από τις αντίστοιχες πηγές (*imdb*, *rotten tomatoes*), κριτικές προϊόντων (*amazon*), κριτικές ταξιδιωτικών προορισμών (*trip advisor*) κ.α.

2. SA in e-learning

Κατά καιρούς έχουν προταθεί διάφοροι ορισμοί ώστε να αποδοθεί επαρκέστερα ο ορισμός του e-learning. Η εκπαίδευση από απόσταση είναι μία διαδικασία μάθησης στην οποία δεν απαιτείται ο εκπαιδευτής και οι εκπαιδευόμενοι να βρίσκονται στον ίδιο τόπο. Ο εκπαιδευτής επικοινωνεί με τους εκπαιδευόμενους με κάποιο μέσο αμφίδρομης επικοινωνίας σύγχρονης ή ασύγχρονης. Η εκπαίδευση αυτή ονομάζεται και τηλε- εκπαίδευση. Υπάρχουν πολλές μορφές εκπαίδευσης από απόσταση. Κάποιες μορφές κάνουν προσομοίωση της διδασκαλίας που γίνεται μέσα στην τάξη με πλήρη επικοινωνία καθηγητών και μαθητών σε πραγματικό χρόνο, ενώ άλλες μορφές υποστηρίζουν την ανεξάρτητη μάθηση που κατευθύνεται από τον εκπαιδευόμενο. Με τη χρήση των υπηρεσιών που προσφέρουν τα περιβάλλοντα ηλεκτρονικής εκμάθησης, κάθε εκπαιδευόμενος θα μπορεί από το χώρο του να παρακολουθεί τη διδασκαλία ενός θέματος όποτε θέλει με το ρυθμό που κρίνει αποδοτικό επιλέγοντας ή επαναλαμβάνοντας τμήματα της και έχοντας ταυτόχρονα πρόσβαση σε μια τεράστια ποικιλία επικουρικού και συμπληρωματικού υλικού, όπως π.χ. ψηφιακές βιβλιοθήκες, ηλεκτρονικά εργαστήρια, εκπαιδευτικό λογισμικό κτλ. E-learning, digital learning, distance learning είναι παρόμοιοι όροι που όλοι τους αναφέρονται σε μια προσπάθεια για μοντέρνα εκπαίδευση που αποσκοπεί κυρίως στο να εκμεταλλευθεί την βασισμένη στο Web τεχνολογία. (Kechaou, 2011).

Η έννοια του e-learning είναι αρκετά γενική και περιλαμβάνει οποιαδήποτε μορφή εκπαίδευσης χρησιμοποιεί τους πόρους του δικτύου ή γενικότερα τις δυνατότητες των ηλεκτρονικών υπολογιστών. Θα μπορούσε να οριστεί ως η βασισμένη στην τεχνολογία μάθηση στην οποία τα υλικά της μάθησης μεταφέρονται ηλεκτρονικά σε απομακρυσμένους χρήστες μέσω ενός δικτύου υπολογιστών. Ένας άλλος σχετικός ορισμός αναφέρει το elearning σαν την βασισμένη στην τεχνολογία εκπαίδευση που συμπεριλαμβάνει την εκπαίδευση που βασίζεται στο Web και την εκπαίδευση που παίρνουμε με την βοήθεια των υπολογιστών. Εκπαίδευση εξ αποστάσεως (Distance Education) ορίζεται, σύμφωνα με την U.S. Distance Learning Association (2006), η απόκτηση γνώσεων και ικανοτήτων με έμμεση πληροφόρηση και καθοδήγηση, που περιλαμβάνει όλες τις τεχνολογίες και άλλες μορφές εκμάθησης από απόσταση. Η Εκπαίδευση εξ αποστάσεως μπορεί να οριστεί ως μια διαδικασία όπου ο διδασκόμενος βρίσκεται σε φυσική απόσταση από τον διδάσκοντα και τον εκπαιδευτικό φορέα, και η τεχνολογία με διάφορους τρόπους διαμεσολαβεί για την εκπαιδευτική διαδικασία (Kechaou, 2011).

Ο όρος Ηλεκτρονική Μάθηση(e- Learning)πρόσφατα προτάθηκε να αντικατασταθεί από τον όρο Τεχνολογικά Ενισχυμένη Μάθηση (Technology Enhanced Learning – TEL) (Browne-Hewitt- Jenkins- Walker,2008).

Το e-learning αποτελεί μια από τις πιο σύγχρονες τηλεματικές εφαρμογές, η οποία χρηματοδοτείται από την Ευρωπαϊκή Ένωση μέσω αρκετών προγραμμάτων. Στόχος της είναι η εκπαίδευση από απόσταση σε εκπαιδευτικά ιδρύματα ,φορείς επιχειρήσεις, άτομα με ειδικές ανάγκες, προβληματικές γεωγραφικές περιοχές από άποψη πρόσβασης κλπ. Έχουν καθοριστεί τρεις διαφορετικές μορφές τηλε-εκπαίδευσης :

i. Η τηλε-εκπαίδευση σε εξατομικευμένο ρυθμό Σε αυτή την περίπτωση προσφέρονται στον εκπαιδευόμενο συνδυασμός εκπαιδευτικών υλικών (βιβλία, αναφορές στο δίκτυο, μαγνητοσκοπημένα μαθήματα, σημειώσεις, προγράμματα εκμάθησης βασισμένα σε υπολογιστή κτλ), συνήθως χωρισμένα σε ενότητες (μαθήματα), τα οποία χρησιμοποιεί με το δικό του ρυθμό, αποφασίζει δηλαδή ο ίδιος πότε και που θα τα χρησιμοποιήσει. Δεν υπάρχει επικοινωνία με τον διδάσκοντα ή με άλλους μαθητές.

ii. Η ασύγχρονη τηλε-εκπαίδευση Η περίπτωση αυτή μοιάζει αρκετά με την προηγούμενη. Παρέχεται στους συμμετέχοντες η δυνατότητα να εργαστούν με το υλικό προς διδασκαλία

οπουδήποτε και οποτεδήποτε έχοντας όμως παράλληλα δυνατότητα ασύγχρονης επικοινωνίας με τους υπόλοιπους συμμετέχοντες και με τον εκπαιδευτή. Το υλικό διδασκαλίας δεν είναι απαραίτητο να έχει δοθεί όλο από την έναρξη του μαθήματος αλλά μπορεί να προσφέρεται τους εκπαιδευόμενους σταδιακά. Ο ρυθμός διεξαγωγής καθορίζεται από τον εκπαιδευτή σε συνεργασία πάντα με τους εκπαιδευόμενους.

iii. Η σύγχρονη τηλεεκπαίδευση Εδώ το μάθημα γίνεται κανονικά αλλά οι μαθητές και ο καθηγητής μπορούν να βρίσκονται σε διαφορετικό τόπο ο καθένας και χρησιμοποιώντας τεχνολογίες τηλεδιάσκεψης να βρίσκονται όλοι σε μία εικονική αίθουσα διδασκαλίας. Η διεξαγωγή του μαθήματος γίνεται με τέτοιο τρόπο ώστε να προσφέρει τις ίδιες ή και παραπάνω δυνατότητες με αυτές που προσφέρονται σε μία κανονική αίθουσα.

Η ολοένα αυξανόμενη διείσδυση των ΤΠΕ (Τεχνολογιών Πληροφορίας και Επικοινωνίας) στην καθημερινή ζωή μας οδηγεί σε 'νέους δρόμους' και τάσεις που επηρεάζουν την παροχή μαθησιακού περιεχομένου μέσω του διαδικτύου. Μια από τις τελευταίες τάσεις είναι τα MOOC (Massive Open Online Courses) ή Μαζικά Ελεύθερα Διαδικτυακά Μαθήματα.

Σύμφωνα με το λεξικό της Οξφόρδης (Oxford Dictionaries, 2013) τα MOOC είναι ένα πρόγραμμα σπουδών που διατίθεται μέσω του διαδικτύου δίχως χρέωση, σ' ένα πολύ μεγάλο αριθμό ανθρώπων. Όποιος αποφασίσει να παρακολουθήσει ένα MOOC συνδέεται στον κατάλληλο δικτυακό τόπο δηλώνοντας απλά την πρόθεση συμμετοχής του. Το ακρωνύμιο του ονόματος MOOC αποπνέει άλλωστε την αρχική ιδέα της σύλληψής τους: 'Massive' (μαζικά) επειδή έχουν σχεδιαστεί ώστε να επιτρέπουν την εγγραφή δεκάδων χιλιάδων εκπαιδευομένων, 'Open' (ελεύθερα), διότι, θεωρητικά, ο καθένας με μια σύνδεση στο διαδίκτυο μπορεί να εγγραφεί στην ελεύθερη σειρά μαθημάτων, 'Online' (διαδικτυακά), διότι ένα σημαντικό μέρος ή αν όχι όλη η αλληλεπίδραση λαμβάνει χώρα διαδικτυακά μέσω ομάδων συζητήσεων, wikis ή με την παρακολούθηση σύντομων βίντεο από τους εκπαιδευόμενους και, τέλος, δεν παύουν να είναι 'Courses' (μαθήματα), διότι έχουν συγκεκριμένες ημερομηνίες έναρξης και λήξης, αξιολόγηση των εκπαιδευομένων, κουίζ και εξετάσεις (Chien- wen Shen et al. 2015)

Τα MOOC θεωρούνται ως η πλέον πρόσφατη εξέλιξη στην εξ αποστάσεως εκπαίδευση. Ουσιαστικά θεωρούνται ως μια εξέλιξη των πλατφόρμων υποβοήθησης των 'παραδοσιακών' μαθημάτων (π.χ. Blackboard, Moodle, κ.ά.) με τη διαφορά ότι επιτρέπουν την ανοικτή πρόσβαση και τη δυνατότητα υποστήριξης πολύ μεγάλου αριθμού εκπαιδευομένων. Τα MOOC μετασχηματίζουν το χώρο της τριτοβάθμιας εκπαίδευσης αλλάζοντας τη σπουδαιότητα και την εμβέλειά των πανεπιστημιακών ιδρυμάτων. Για παράδειγμα, το 2012 στο MOOC που διοργάνωσε το πανεπιστήμιο του Stanford με τίτλο 'Artificial Intelligence' εγγράφηκαν 160.000 φοιτητές από 190 χώρες με 20.000 από αυτούς να το ολοκληρώνουν (Breslow et al., 2013).

Σύμφωνα με τους Chien-wen Shen et al (2015), τα MOOC φιλοδοξούν να παρέχουν δωρεάν πρόσβαση σε μαθήματα και τεχνολογίες αιχμής, οδηγώντας βραχυπρόθεσμα σε μείωση του κόστους της εκπαίδευσης πανεπιστημιακού επιπέδου και, ενδεχομένως, μακροπρόθεσμα στην αναδιοργάνωση των υπάρχοντων μοντέλων της τριτοβάθμιας εκπαίδευσης. Η παραπάνω συλλογιστική έχει ενθαρρύνει κορυφαία πανεπιστήμια να 'ανεβάσουν' τα μαθήματά, τους στο διαδίκτυο μέσω της δημιουργίας ανοικτών πλατφορμών μάθησης, όπως την edX. Επίσης, νεοσύστατα επιχειρηματικά σχήματα, όπως η Coursera και η Udacity, έχουν ξεκινήσει σε συνεργασία με έγκριτα πανεπιστήμια να προσφέρουν δωρεάν διαδικτυακά μαθήματα ή χρεώνοντας ένα μικρό ποσό στην περίπτωση που κάποιος εκπαιδευόμενος επιθυμεί να λάβει μέρος σε μια διαδικασία πιστοποίησης των γνώσεών του.

Είναι γνωστό πως εταιρείες κολοσσοί, όπως η Pearson και η Google, σχεδιάζουν να διεισδύσουν στο χώρο της τριτοβάθμιας εκπαίδευσης ως παγκόσμιοι παίκτες και πιθανότατα θα υιοθετήσουν μια MOOC προσέγγιση ως μέρος των σχεδίων τους. Στην Ευρώπη, μια νέα εταιρεία, η Futurelearn,

δημιουργήθηκε στο Ηνωμένο Βασίλειο από το Open University προκειμένου να συγκεντρώσει και να παρέχει δωρεάν, ανοικτά, διαδικτυακά μαθήματα από τα κορυφαία πανεπιστήμια του Ηνωμένου Βασιλείου (Keerthi et al, 2014).

Ο Hanley (2013) επισημαίνει ότι οι αισιόδοξοι βλέπουν σημαντικές μελλοντικές συνέργειες στον ανοικτό διαδικτυακό χαρακτήρα των MOOC θεωρώντας τα ως το όχημα για την επέκταση της πρόσβασης σε οικονομικά προσιτή μάθηση, την ανάπτυξη νέων παιδαγωγικών μεθόδων και τη στήριξη μιας παγκόσμιας κοινότητας μάθησης για την από κοινού δημιουργία της γνώσης. Οι απαισιόδοξοι θεωρούν τα MOOC ως ένα είδος δούρειου ίππου, ο οποίος θα εισβάλει στο χώρο της Τριτοβάθμιας Εκπαίδευσης προκειμένου να καταστρέψει ιδρύματα, να παροπλίσει σχολές και να υποβαθμίσει τη μάθηση μετατρέποντάς την σε μια εμπειρία τύπου 'fast food', δηλαδή μιας άμεσης αλλά κενής περιεχομένου μαθησιακής εμπειρίας (Keerthi et al, 2014).

Στην πραγματικότητα, τα MOOC αποτελούν μια πάρα πολύ νέα εκπαιδευτική προσέγγιση για να υπάρξουν επαρκή επιστημονικά ευρήματα σχετικά με την αξία, το κόστος και τους κινδύνους που απορρέουν από τη διάδοσή τους. Τα πιθανά οφέλη ή οι απειλές για την ακαδημαϊκή ποιότητα, τα μαθησιακά αποτελέσματα, τη θεσμική ακεραιότητα των ακαδημαϊκών ιδρυμάτων δεν είναι ακόμη γνωστά (Yuan & Powell, 2013).

Από τη βιβλιογραφία που χρησιμοποιήθηκε για την παρούσα μελέτη διαπιστώθηκε πως τα πιο διαδεδομένα μέσα κοινωνικής δικτύωσης στην ηλεκτρονική εκπαίδευση είναι το Facebook και το Twitter.

Αρχικά κάνοντας λόγο για το Facebook, φαίνεται ότι ένα από τα πιο σημαντικά οφέλη της χρήσης του για τους εκπαιδευόμενους, είναι ότι διαπιστώνεται ενεργητική συμμετοχή και αλληλεπίδραση (Ortigosa et al, 2013), αναπτύσσεται η θετική συμπεριφορά κατά τη διάρκεια της μάθησης και βελτιώνεται η ποιότητα μάθησης (Kirschner & Karpinski, 2010). Αναπτύσσεται η αλληλεπίδραση των συμμετεχόντων και η κριτική σκέψη (Lampe et al., 2008) και οι εκπαιδευόμενοι δημιουργούν το δικό τους ρυθμό μάθησης, χρησιμοποιώντας ειδικές εφαρμογές και εργαλεία (Hew, 2011). Παράλληλα, καλλιεργείται η ανοχή και ο σεβασμός στη διαφορετικότητα. (Young & QuanHaase, 2009), ενισχύεται η αυτοπεποίθηση και η αυτοεκτίμηση (Bosch, 2009), ενώ δημιουργούνται και θετικές προοπτικές στην αντιμετώπιση των προκλήσεων που αντιμετωπίζουν οι μαθητές. (Ortigosa et al, 2013)

Οι χρήστες του Facebook, διαμοιράζονται αρκετά από τα ποιοτικά στοιχεία των εκπαιδευτικών μεθόδων που χρησιμοποιούνται στην παραδοσιακή τάξη, τα οποία δημιουργούνται από την αλληλεπίδραση των χρηστών μεταξύ τους, μέσα σε ένα καλά δομημένο κοινωνικό πλαίσιο μάθησης (Ortigosa et al, 2013).

Τα οφέλη της χρήσης του Facebook για τους εκπαιδευτές, είναι ότι εφαρμόζονται καινοτόμες παιδαγωγικές μέθοδοι, με σκοπό την αύξηση ενδιαφέροντος των εκπαιδευόμενων και αναπτύσσονται γνωστικές ικανότητες και δεξιότητες, με σκοπό την αποτελεσματικότητα της μαθησιακής διαδικασίας. Επειδή το Facebook, προσφέρει το κατάλληλο κοινωνικό, οργανωτικό και εννοιολογικό πλαίσιο (Hew, 2011), αποδέχεται τους συμμετέχοντες σαν αλληλεπιδρώντες χρήστες και τους εκπαιδευτές σαν καθοδηγητές (Schwartz, 2009).

Ωστόσο, παρά τις ευκολίες που παρέχουν τα κοινωνικά δίκτυα στην εκπαίδευση, οι εκπαιδευτές πρέπει να καταβάλλουν ιδιαίτερες προσπάθειες για την προώθηση της ενεργητικής μάθησης και της αποτελεσματικότητας των on-line ομάδων μάθησης. Τα μειονεκτήματα που δημιουργούνται από τη χρήση των κοινωνικών δικτύων στην εκπαίδευση, είναι κυρίως, η ανακρίβεια και η παροχή μη έγκυρων πληροφοριών (De Villiers, 2010). Η έλλειψη εμπιστοσύνης στις δημοσιεύσεις, οδηγεί σε αισθήματα ανασφάλειας. Ο Hargadon (2009) τονίζει, ότι στα κοινωνικά δίκτυα, συχνά εμφανίζονται φαινόμενα όπως κακόβουλα σχόλια και ναρκισσισμός. Σύμφωνα με τους Liccardi et al. (2007), η παραβίαση προσωπικών πληροφοριών και η δημιουργία ψευδών λογαριασμών, οδηγούν σε ανασφάλεια και κατά συνέπεια, σε περιορισμό της χρήσης των κοινωνικών δικτύων.

Αναφορικά με το Twitter, θα λέγαμε ότι είναι ένας από τους πλέον δημοφιλείς ιστότοπους κοινωνικής δικτύωσης που επιτρέπει στους χρήστες να κοινοποιήσουν μηνύματα μεγέθους το πολύ 140 χαρακτήρων που καλούνται tweets. Δημιουργήθηκε το Μάρτιο του 2006 και μέσα σε μικρό χρονικό διάστημα προσέλκυσε πολύ μεγάλο αριθμό χρηστών. Πλέον εξυπηρετεί περισσότερους από 310 εκατομμύρια ενεργούς χρήστες το μήνα.

Τα βασικά χαρακτηριστικά των μηνυμάτων που ανταλλάσσονται μέσω Twitter είναι Μικρό μέγεθος και περιεκτικότητα, το πολύ 140 χαρακτήρες.

Ανεπίσημος λόγος, πολλές φορές χωρίς έμφαση στη σύνταξη και τη γραμματική ορθότητα.

Πολύ συχνά περιέχουν συντομεύσεις και ακρωνύμια.

Επιμηκυμένες λέξεις, επαναλήψεις σημείων στίξης και λέξεις γραμμένες μόνο με κεφαλαία, όλα προς απόδοση έμφασης.

Ειδικά tokens όπως hashtags, usernames, URLs, retweets και emoticons.

Τα hashtags χαρακτηρίζονται από τον χαρακτήρα #, με τον οποίο πάντα ξεκινάνε, και την απουσία κενών. Μπορεί να δηλώνουν το θέμα του tweet, συναισθηματική κατάσταση ή και οτιδήποτε άλλο. Συνήθως χρησιμοποιούνται για την επισήμανση κάποιας τάσης (trend) μεταξύ των χρηστών. Το σώμα των hashtags μπορεί να περιέχει μόνο πεζούς χαρακτήρες αλλά επίσης συνηθίζεται να περιέχει κεφαλαία για τον διαχωρισμό των λέξεων καθώς δεν επιτρέπεται η χρήση κενών.

Τα usernames ξεκινάνε με τον χαρακτήρα @ και επισημαίνουν κάποιον χρήστη του δικτύου. Το token RT δηλώνει retweet δηλαδή κοινοποίηση του μηνύματος άλλου χρήστη. Τα URLs είναι ηλεκτρονικές διευθύνσεις που επισημαίνονται σε μηνύματα. Ξεκινάνε σχεδόν πάντα με http://, https:// ή www. (Ravicharidran et al, 2014)

Τα emoticons είναι strings που προσομοιάζουν ανθρώπινες εκφράσεις με χαρακτήρες για να δηλώσουν συναισθηματική κατάσταση όπως χαρά, λύπη, θυμό, έκπληξη και γέλιο. Η πληροφορία που μεταφέρουν είναι πολύτιμη για το πρόβλημα της ανάλυσης συναισθήματος. (Sahayak et al, 2015)

Ωστόσο, σε δεδομένα Twitter είναι δυνατόν να αυτοματοποιηθεί η διαδικασία χαρακτηρισμού της συναισθηματικής πολικότητας των tweets με αυτόματο χαρακτηρισμό των tweets με θετικά emoticons ως θετικά και των αντίστοιχων με αρνητικά emoticons ως αρνητικά. Μία τέτοια μέθοδος καλείται noisy labeling (ή distant supervision) καθώς στις επιθυμητές εξόδους ενυπάρχει θόρυβος. Tweets με θετικά emoticons δεν είναι απαραίτητα θετικά και αντίστροφα. Η συνολική συναισθηματική πολικότητα βέβαια ενός συνόλου λέξεων, ειδικά ενός τόσο μικρού όσο ένα tweet είναι πολλές φορές δύσκολο να χαρακτηριστεί απόλυτα.

Οι annotators πολλές φορές διαφωνούν σε μικρό ποσοστό ακόμα και στον χαρακτηρισμό πολύ συγκεκριμένου περιεχομένου, όπως αντικείμενα σε εικόνες, συνεπώς σε κάτι πιο αφηρημένο όπως το συναίσθημα σε ένα tweet τα ποσοστά διαφωνίας είναι αρκετά μεγαλύτερα. Από τα παραπάνω προκύπτει ότι η συμφωνία των annotators είναι ένα στοιχείο που πρέπει να λαμβάνεται υπόψη στην αξιολόγηση ενός sentiment analysis αλγορίθμου. Ποσοστό επιτυχίας 98% σε ένα πρόβλημα αναγνώρισης αντικειμένων, όταν το αντίστοιχο ποσοστό ενός ανθρώπου είναι 99% δείχνει ότι ο αλγόριθμος λειτουργεί ικανοποιητικά. Αντίθετα ποσοστό επιτυχίας 98% σε ένα πρόβλημα ανάλυσης συναισθήματος στο οποίο ο άνθρωπος συμφωνεί κατά 90% με τους χαρακτηρισμούς των annotators δηλώνει ότι ο αλγόριθμος ταξινόμησης λειτουργεί σωστά αλλά το πρόβλημα δεν επιλύεται ικανοποιητικά. Σε κάθε περίπτωση ο χαρακτηρισμός των δεδομένων πρέπει να γίνεται από πολλούς annotators για να εξασφαλίζεται η ποιότητα των δεδομένων και να εκπαιδεύονται συστήματα που συλλαμβάνουν την πραγματική διάσταση του προβλήματος. (Harris et al, 2014)

Παρατηρείται λοιπόν πως τα μέσα κοινωνικής δικτύωσης είναι ολοένα και πιο διαδεδομένη μορφή ηλεκτρονικής μάθησης. Ωστόσο, προκειμένου να γίνει κάποια συλλογή δεδομένων για τον κάθε φοιτητή χρησιμοποιούνται συστήματα αλγορίθμων και διανυσμάτων, (Bayes, SVM) ,όπως

προαναφέρθηκε, προκειμένου να καταγραφούν οι αντιδράσεις των εκπαιδευομένων απέναντι στη μαθησιακή κατάσταση. Στο παρακάτω κεφάλαιο θα γίνει μια σύγκριση αυτών των συστημάτων, παρουσιάζοντας τα θετικά και τα αρνητικά τους χαρακτηριστικά.

3. Comparative Analysis & Discussion

Στο συγκεκριμένο κεφάλαιο θα προσπαθήσουμε να ορίσουμε ορισμένα κριτήρια με τα οποία θα αναλύσουμε τις προσεγγίσεις ανάλυσης συναισθήματος που αναφέρθηκαν στις προηγούμενες ενότητες.

Έτσι λοιπόν, βασικά κριτήρια είναι τα εξής:

Εγκυρότητα (Validity) : Η εγκυρότητα αναφέρεται στο εάν μια μέτρηση αντανακλά την αντικείμενο που πρόκειται να μετρηθεί. Μπορεί το εργαλείο ανάλυσης συναισθήματος να δώσει αποτελέσματα για αυτό που προορίζεται να μετρηθεί ή μήπως δίνει αποτελέσματα για κάτι άλλο;

Αξιοπιστία (Reliability) Αξιοπιστία παρατήρησης σημαίνει ότι τα ίδια δεδομένα θα μετριοούνται κάθε φορά σε μια επαναλαμβανόμενη παρατήρηση. Μπορεί το εργαλείο ανάλυσης συναισθήματος να δώσει τα ίδια αποτελέσματα όταν η αναζήτηση επαναλαμβάνεται;

Ακρίβεια (Precision) Η Ακρίβεια αναφέρεται στο πόσο ακριβής μέτρηση είναι. Είναι τα αποτελέσματα ακριβή ή είναι ασαφείς; Γενικός κανόνας εδώ είναι ότι πιο ακριβής είναι καλύτερο από ό, τι λιγότερο ακριβείς.

Σχετικότητα (Relevancy) Η Σχετικότητα είναι για το αν η παραγόμενη γνώση έχει σημασία για την επίλυση του προβλήματος πρακτικά. Μπορεί αυτό το εργαλείο ανάλυσης συναισθήματος να παρέχει πληροφορίες που είναι σχετικές με το πρόβλημα που προσπαθούμε να λύσουμε;

Η διαφάνεια του συλλογισμού (Transparency of reasoning) Η διαφάνεια του συλλογισμού είναι για το αν το εργαλείο ανάλυσης συναισθήματος δείχνει το πώς

ήρθε στα συμπεράσματά του. Μήπως το εργαλείο σας παρέχει τη δυνατότητα να δείξει τον τρόπο με τον οποίο οδηγούμαστε από τις πληροφορίες στα συμπεράσματα ή μήπως δίνει μόνο τα συμπεράσματα;

Λογική συνέπεια (Logical consistency) Η λογική συνέπεια αναφέρεται στη αιτιότητα της κάθε παραγόμενης γνώσης. Είναι ο συλλογισμός λογικός; Ή μήπως τα συμπεράσματα της ανάλυσης δεν έχουν κανένα νόημα;

Ανάλυση του σκοπού(Detail of reasoning) Η Ανάλυση του σκοπού αναφέρεται στο αν η αιτιολόγηση από το εργαλείο ανάλυσης συναισθήματος είναι λεπτομερής ή όχι. Είναι σημαντικό για το εργαλείο ανάλυσης συναισθήματος να χρησιμοποιεί τις λεπτομερείς αιτιολογίες ως αιτιολογία διότι αν είναι πάρα πολύ απλό, μπορεί να οδηγήσει σε λάθος συμπεράσματα

Κατανόηση για το ευρύ κοινό (Understandability for laymen) Η Κατανόηση για το ευρύ κοινό αναφέρεται στο εάν τα αποτελέσματα δίνονται με έναν τρόπο που είναι κατανοητό για μια απλή γλώσσα. Τα αποτελέσματα παρουσιάζονται με μη-τεχνικούς όρους και κατανοητό τρόπο;

Ωφελιμότητα των πληροφοριών (Usability of information)

Η ωφελιμότητα των πληροφοριών αναφέρεται στο κατά πόσον το εργαλείο ανάλυσης συναισθήματος παράγει χρήσιμες πληροφορίες. Μήπως το εργαλείο παρέχει άμεσα ωφέλιμες πληροφορίες;

Κατηγορίες(Categories) Οι Κατηγορίες αναφέρονται στον αριθμό των κατηγοριών που παρέχει το εργαλείο ανάλυσης συναισθήματος. Το εργαλείο ανάλυσης συναισθήματος πρέπει να εξετάζει αρκετές κατηγορίες

Συνάφεια των κατηγοριών (Relevancy of categories)

Συνάφεια των κατηγοριών αναφέρεται στο κατά πόσον οι κατηγορίες που παρέχονται από το εργαλείο ανάλυσης συναισθήματος είναι σχετικές.

Πληρότητα των κατηγοριών(Completeness of categories)

Η πληρότητα των κατηγοριών αναφέρεται στο αν οι κατηγορίες που παρέχει το εργαλείο ανάλυσης συναισθήματος είναι πλήρεις

Διαφάνεια για την αντιμετώπιση των προκαταλήψεων (Transparency of dealing with biases)

Η Διαφάνεια ασχολείται με προκαταλήψεις σχετικές με το κατά πόσον το εργαλείο ανάλυσης συναισθήματος διευκρινίζει τον τρόπο με τον οποίο διαπραγματεύεται με προκαταλήψεις;

Περιορισμός της προκατάληψης (Elimination of bias)

Ο Περιορισμός της μεροληψίας αναφέρεται στο κατά πόσον το εργαλείο ανάλυσης συναισθήματος λαμβάνει υπόψη τις προκαταλήψεις. Έχει λάβει δημογραφική προκατάληψη, εκδηλώσεων και η μεροληψία χειραγώγησης υπόψη κατά την ανάλυση των δεδομένων;

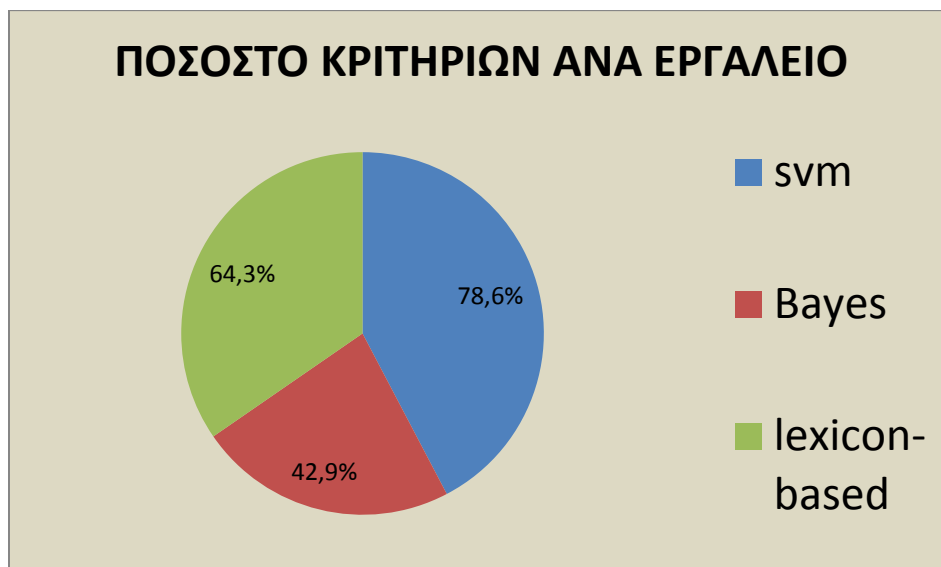
Στον Πίνακα που ακολουθεί παρουσιάζονται τα κριτήρια σχετικά με τα εργαλεία προσέγγισης για την ανάλυση συναισθήματος και κατά πόσο τα εργαλεία αυτά ανταποκρίνονται στα κριτήρια που θέτουμε.

Criteria	svm	Bayer	lexicon-based
Validity	χ	χ	χ
Reliability	χ	χ	χ
Precision	χ		
Causality	χ	χ	
Detail of reasoning			
Transparency of reasoning			χ
Understandability for laymen			χ
Relevancy for problem solving	χ		χ
Usability of information	χ		χ
Transparency of dealing with biases	χ		
Elimination of biases	χ	χ	χ
Categories	χ	χ	χ
Relevancy	χ	χ	χ
Completeness	χ		

Πίνακας 1. Αξιολόγηση εργαλείων ανάλυσης συναισθήματος με βάση ορισμένα κριτήρια.

Με βάση τον παραπάνω πίνακα παρατηρούμε πως ορισμένα συστήματα σχετίζονται με κάποια από τα κριτήρια που ορίσαμε και κάποια όχι. Για παράδειγμα οι αλγόριθμοι SVM και το σύστημα Bayes δίνουν μια εγκυρότητα και αξιοπιστία, σχετικότητα και διαφάνεια ως προς τις πληροφορίες που ζητάμε από την ανάλυση συναισθήματος αλλά δεν είναι απόλυτα κατανοητοί και δεν προσφέρουν πολλές λεπτομέρειες. Από την άλλη πλευρά τα lexicon based είναι περισσότερο εύχρηστα, παρέχουν αξιοπιστία και διαφάνεια αλλά δέχονται ορισμένους περιορισμούς και δεν είναι τόσο αναλυτικοί ως προς την έκφραση συναισθημάτων. Με βάση και το γράφημα που ακολουθεί (βλ. γράφημα 1) παρατηρούμε ότι οι αλγόριθμοι SVM καλύπτουν τα περισσότερα κριτήρια από αυτά που θέσαμε σε ποσοστό 78%. Δεύτερα έρχονται τα lexicon based με 64% και τελευταίο το σύστημα Bayes με 42,9%. Από τα 14 κριτήρια ωστόσο που θέσαμε πολλά είναι κοινά και στα τρία εργαλεία όπως για παράδειγμα η Reliability, Validity, Elimination of biases, Categories, Relevancy. Αξίζει επιπλέον να σημειωθεί πως οι αλγόριθμοι SVM είναι το μονό από τα τρία εργαλεία που παρέχει διαφάνεια για την αντιμετώπιση των προκαταλήψεων (Transparency of dealing with biases, ακρίβεια (Precision) αλλά και Πληρότητα των κατηγοριών (Completeness of categories). Επίσης φαίνεται πως τα lexicon based είναι τα μόνα εργαλεία από τα τρία που έχει διαφάνεια του συλλογισμού (Transparency of reasoning και Κατανόηση για το ευρύ κοινό (Understandability for laymen ενώ τέλος κανένα από τα τρία εργαλεία δεν διαθέτει το κριτήριο της Ανάλυσης του σκοπού (Detail of reasoning) που σημαίνει ότι κανένα από τα τρία εργαλεία δεν παρέχουν όλες τις λεπτομέρειες που χρειάζονται ως προς την ανάλυση συναισθημάτων προκειμένου να προβούν σε απόλυτα συμπεράσματα. Με βάση τα παραπάνω, όλα τα εργαλεία αν και τηρούν πολλά από τα σημαντικά κριτήρια που θέσαμε, κανένα δεν μπορεί να οδηγήσει σε συμπεράσματα με απόλυτη ακρίβεια.

ΕΡΓΑΛΕΙΑ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ	ΠΟΣΟΣΤΟ ΚΡΙΤΗΡΙΩΝ
SVM	78,6%
Bayes	42,9%
Lexicon-based	64,3%



Γράφημα 1. Ποσοστό κριτηρίων ανά εργαλείο

4. Conclusion & future work

Με βάση τα παραπάνω είδαμε αρχικά ότι από τις πρώτες μελέτες στην περιοχή της Ανάλυσης Συναισθήματος (Peter D Turney. 2002) είχε ήδη γίνει αντιληπτό ότι η κατηγοριοποίηση συναισθήματος (sentiment classification) διαφέρει από το κλασικό πρόβλημα κατηγοριοποίησης 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”*(Fabrizio Sebastiani 2002)

Οι κατηγορίες καθορίζονται με βάση τα θέματα στόχους του εκάστοτε προβλήματος. Επομένως, διαφορετικά προβλήματα ταξινόμησης κειμένου βασίζονται σε διαφορετικά σύνολα κατηγοριών. Το πλήθος των κατηγοριών σε ένα σύνολο ποικίλει : μπορεί να εκτείνεται από ένα μικρό σύνολο δύο μόνο κατηγοριών έως σύνολα με δεκάδες κατηγορίες π.χ. οι κατηγορίες που απαιτούνται για την ταξινόμηση ενός άρθρου εφημερίδας με βάση τη θεματολογία που καλύπτει. Παράλληλα, ανάλογα με το πρόβλημα και το σύνολο κατηγοριών, ένα κείμενο μπορεί να ανήκει σε μία ή περισσότερες επικαλυπτόμενες κατηγορίες π.χ. ένα άρθρο να αντιστοιχηθεί με τις κατηγορίες “πολιτική”, “οικονομία” και “επικαιρότητα”.(Bo Pang and Lillian Lee. 2008)

Αντίθετα, η Ανάλυση Συναισθήματος αναφέρεται σε ένα μικρό σύνολο κατηγοριών (π.χ. θετικό, αρνητικό, ουδέτερο - “1 αστέρι”, ...,“5 αστέρια”). Επειδή επικεντρώνεται στην κατάταξη ενός κειμένου ως προς την πολικότητα του, οι κατηγορίες είναι ανεξάρτητες της θεματολογίας του προβλήματος και μεταξύ τους αμοιβαία αποκλειόμενες.

Το εννοιολογικό πλαίσιο στο οποίο κινείται η Ανάλυση Συναισθήματος εξασφαλίζει την εφαρμογή των τεχνικών και μεθόδων της σε ένα μεγάλο εύρος θεμάτων χωρίς ιδιαίτερες τροποποιήσεις, επιτυγχάνοντας αρκετά ικανοποιητικά ποσοστά ακρίβειας. Στηριζόμενοι στην ανεξαρτησία του προβλήματος από τη θεματολογία, θα μπορούσαμε να ισχυριστούμε ότι η πολικότητα ενός κειμένου προκύπτει από την πολικότητα των μεμονωμένων λέξεων από τις οποίες απαρτίζεται. Συνεπώς, αναγνωρίζοντας ένα συγκεκριμένο σύνολο λέξεων-κλειδιών (keywords) θα μπορούσαμε να προσδιορίσουμε τη συνολική πολικότητα της άποψης που εκφράζεται στο κείμενο.

Η παραπάνω διαδικασία είναι μία από τις πρώτες μεθόδους που χρησιμοποιήθηκαν και υιοθετεί μία από τις πιο δημοφιλείς και αποτελεσματικές τεχνικές της ανίχνευσης θεματολογίας. Ωστόσο, η προσέγγιση μέσω λέξεων κλειδιών στο συγκεκριμένο πρόβλημα δεν εμφανίζει υψηλά ποσοστά ακρίβειας και έχει αποδεχθεί ελλιπής σε ορισμένες περιπτώσεις (“thwarted expectations” (Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002).

Στο σημείο αυτό ανακύπτει το εξής ερώτημα : για ποιο λόγο το πρόβλημα της κατηγοριοποίησης συναισθήματος είναι πιο δύσκολο σε σχέση με την ανίχνευση θεματολογίας, αν λάβουμε υπόψη ότι οι κατηγορίες “θετικό”, “αρνητικό” και “ουδέτερο” είναι εννοιολογικά ξένες μεταξύ τους ;

Μία από τις πιο σημαντικές διαφορές με την κατηγοριοποίηση ως προς τη θεματολογία και τις δυσκολίες στην περιοχή της Ανάλυσης Συναισθήματος είναι ότι “το συναίσθημα/άποψη μπορεί πολλές φορές να εκφραστεί με πιο λεπτό τρόπο χωρίς τη χρήση συναισθηματικά φορτισμένων (θετικά ή αρνητικά) λέξεων με αποτέλεσμα να είναι δύσκολο να αναγνωρισθεί από τους επιμέρους όρους του κειμένου όταν αυτοί εξετάζονται μεμονωμένα” .

Παράλληλα, πέρα από τον προσδιορισμό της πολικότητας όταν απουσιάζουν συναισθηματικά φορτισμένες λέξεις, ιδιαίτερα απαιτητικός είναι και ο διαχωρισμός των υποκειμενικών και αντικειμενικών λέξεων και φράσεων ενός κειμένου. Όπως αναφέρεται από τους Kim και Hong στο

“πολλές φορές ακόμη και άνθρωποι διαφωνούν για το αν μία δήλωση αποτελεί ή όχι άποψη”. (Soo-Min Kim and Eduard Hovy. 2006)

Ένα άλλο ζήτημα που απασχολεί ιδιαίτερα την Ανάλυση Συναισθήματος είναι ο προσδιορισμός του κατόχου - εκφραστή της άποψης (opinion holder) που διατυπώνεται στο κείμενο.

Όπως αναφέρθηκε στην παρούσα εργασία, η γενικότερη αντίληψη της θετικής ή αρνητικής άποψης δεν εξαρτάται άμεσα από το εκάστοτε θέμα συζήτησης. Ωστόσο, το συναίσθημα και η υποκειμενικότητα ενός κειμένου εξαρτώνται από το σημασιολογικό πλαίσιο στο οποίο τοποθετείται. (Bo Pang and Lillian Lee. 2008) Χαρακτηριστικό παράδειγμα: “πήγαινε διάβασε το βιβλίο”. Η πρόταση εκφράζει θετική άποψη όταν αναφέρεται σε κριτική βιβλίου. Η ίδια πρόταση, όμως, εκφράζει εντελώς διαφορετική άποψη όταν χρησιμοποιείται σε κριτική ταινίας. Άλλος ένας παράγοντας που επηρεάζει την πολικότητα είναι η σειρά των λέξεων και φράσεων στο κείμενο (Bo Pang and Lillian Lee.2008). Οι ίδιες λέξεις με διαφορετική σειρά μπορεί να οδηγήσουν σε τελείως διαφορετική συνολική πολικότητα.

Τέλος, στις δυσκολίες που συναντά η Ανάλυση Συναισθήματος πρέπει να συμπεριληφθούν και οι προκλήσεις της ευρύτερης περιοχής της Επεξεργασίας Φυσικής Γλώσσας: αμφισημία, χειρισμός της άρνησης, ειρωνεία και σαρκασμός.

Επιπλέον είδαμε ότι στο χώρο της εκπαίδευσης πλέον , και ειδικότερα στην τριτοβάθμια, χρησιμοποιείται η εξ αποστάσεως διδασκαλία με τη βοήθεια των μέσων κοινωνικής δικτύωσης όπως το Facebook και το Twitter.

Παράλληλα, μέσα από την ανάλυση ορισμένων εργαλείων (αλγόριθμοι SVM., BAYES και μέθοδο lexicon based) ανάλυσης συναισθήματος που κάναμε με βάση ορισμένα κριτήρια, διαπιστώθηκε πως να μην παρέχουν κάποια αξιοπιστία και ωφελιμότητα αλλά υστερούν στην ακρίβεια και στην λεπτομερή ανάλυση όταν αναφέρονται στο Twitter και Facebook.

Η παρούσα μελέτη ασχολήθηκε με ένα μέρος του προβληματισμού για την ανάλυση συναισθήματος στο χώρο της εκπαίδευσης. Τα ζητήματα και ουσιαστικά, μειονεκτήματα που παρουσιάζει ο τομέας ανάλυσης συναισθήματος δίνει το πεδίο για να μπορέσει να γίνει μια περαιτέρω μελέτη πάνω σε αυτό το θέμα, με βάση και τη σημερινή μελέτη.

Βιβλιογραφία

A. Ortigosa, R.M. Carro and J.I. Quiroga, (2013)"Predicting user personality by mining social interactions in Facebook," *Journal of Computer and System Sciences*, vol. 80, pp. 57 – 71

B. Liu, L. Zhang, (2012). A survey of opinion mining and sentiment analysis, *Mining Text Data*, 415-463.

Bo Pang and Lillian Lee. (2008) Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. (2002)Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL- 02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics,.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in

Cambria, E.; Song, Y.; Wang, H. & Howard, N. (2013), 'Semantic Multi-Dimensional Scaling for Open-Domain Sentiment Analysis', *Intelligent Systems, IEEE PP*(99), 1-1.

Cambria, Erik. 2013. "An introduction to concept-level sentiment analysis." *Advances in Soft Computing and Its Applications*. Springer Berlin Heidelberg, 478-483

Chen X. (2014) Mining Social Media Data for Understanding Students Learning Experiences. *ieee transactions on learning technologies*, vol. 7, no. 3,

Chien-wen Shen Chin-Jin Kuo (2015) Learning in massive open online courses: Evidence from social media mining *Computers in Human Behavior* 51 568–577

De Villiers, M.R. (2010). Academic use of a group on Facebook: Initial findings and perceptions. *Proceedings of Informing Science & IT Education Conference 2010* (pp.173-190), Cassino: Italy

Eugenio martínez-cámara, m. Teresa martín-valdivia, I. Alfonso ureña-lópez and Arturo Montejó-Ráez (2014) sentiment analysis in twitter natural language engineering / volume 20 / issue 01 pp 1 – 28

Fabrizio Sebastiani. (2002) .Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47,

Francesco Colace, Massimo De Santo, Luca Greco (2014) safe: a sentiment analysis framework for e-learning, vol 9

Hanley, G. (2013). A Message from the MERLOT Executive Director: MOOCs, MERLOT, and Open Educational

Hargadon, S. (2009.) Illuminate educational networking: The important role Web2.0 will play in education. Social Learning Consultant. Retrieved 8 October 2011, from <http://audio.edtechlive.com/lc/EducationalSocialNetworkingWhitepaper.pdf>, προσπελάστηκε 10/8/2016

Kirschner, P., & Karpinski, A. (2010). Facebook and academic performance. *Computers in Human Behavior*, 26, 1237–1245

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques.

Liccardi, I., Ounnas, A., Pau, R., Massey, E., Kinnumen, P., Lewthwaite, S., Midy, M. A., & Sakar, C. (2007). The role of social networks in students' learning experiences. *ACM SIGCSE Bulletin*, 39(4), 224-237

Peter D Turney. (2002) 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*, pages 417–424. Association for Computational Linguistics,.

Sarah Schrauwen (July 2010) .Machine Learning Approaches to Sentiment Analysis using the Dutch Netlog Corpus. University of Antwerp Services. *MERLOT Journal of Online Learning and Teaching*, 9(2).

Soo-Min Kim and Eduard Hovy. (2006) Identifying and analyzing judgment opinions. In *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, HLT-NAACL* pages 200–207, Stroudsburg, PA, USA,. Association for Computational Linguistics

Steven C. Harris¹, Lanqin Zheng², Vive Kumar¹, Kinshuk¹ (2014) Multi-Dimensional Sentiment Classification in Online Learning Environment, *IEEE Sixth International Conference on Technology for Education the worldwide classroom: Research into edX's first MOOC. Research & Practice in Assessment*, 8, 13-25.

Thomas, Keerthi; Fern´andez, Miriam; Brown, Stuart and Alani, Harith (2014). OUSocial2: a platform for gathering students' feedback from social media. In: *The 13th International Semantic Web Conference (ISWC 2014)*, 19-23 October 2014, Riva Del Garda, Trentino, Italy.

Vohra S.M, Teraiya B, (2013) 'A Comparative study of sentiment analysis Techniques, vol 2, issue 2

Young, A.L., & Quan-Haase, A. (2009). Information revelation and internet privacy concerns on social network sites: A case study of Facebook. *Proceedings of the 4th International Conference on Communities and Technologies* (pp. 265–274). NY: ACM.

wikipedia.org/wiki/Blog, προσπελάστηκε 10/8/2016

www.webz.gr/2007/11/13/web-20-goodies/ προσπελάστηκε 10/8/2016