



ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΕΛΛΑΔΟΣ

M105: Κοινωνική Δικτύωση
Χειμερινό Εξάμηνο 2019-2020

Project: Sentiment Analysis

ΟΜΑΔΑ 1

Σαουλίδης Χάρης, Γιαννόπουλος Αλέξανδρος,
Θοδωρής Γκουντής, Γεωργιάδης Κωνσταντίνος,
Βασιλειάδης Παύλος

BHMA 1 INPUT TWEETS

Χρησιμοποιήθηκε η Clojure για να αναπτυχθεί μια εφαρμογή η οποία θα λαμβάνει τα tweets και θα τα αποθηκεύει σε μια sql database. Τα tweets που χρησιμοποιήθηκαν αφορούν 182k tweets από χρήστες για online games. Αφορούν κυρίως παιχνίδια σε πραγματικό χρόνο (Real Time).

BHMA 2 PRE PROCESSING

Κατά την διάρκεια του Pre Processing αφαιρέθηκαν τα κενά από τα tweets για να μπορέσουν να γίνουν import στο SentiStrength και και το επόμενο βήμα της εξαγωγής του Sentiment. Έγινε επίσης το απαραίτητο tokenization και η απαλοιφή των html tags.

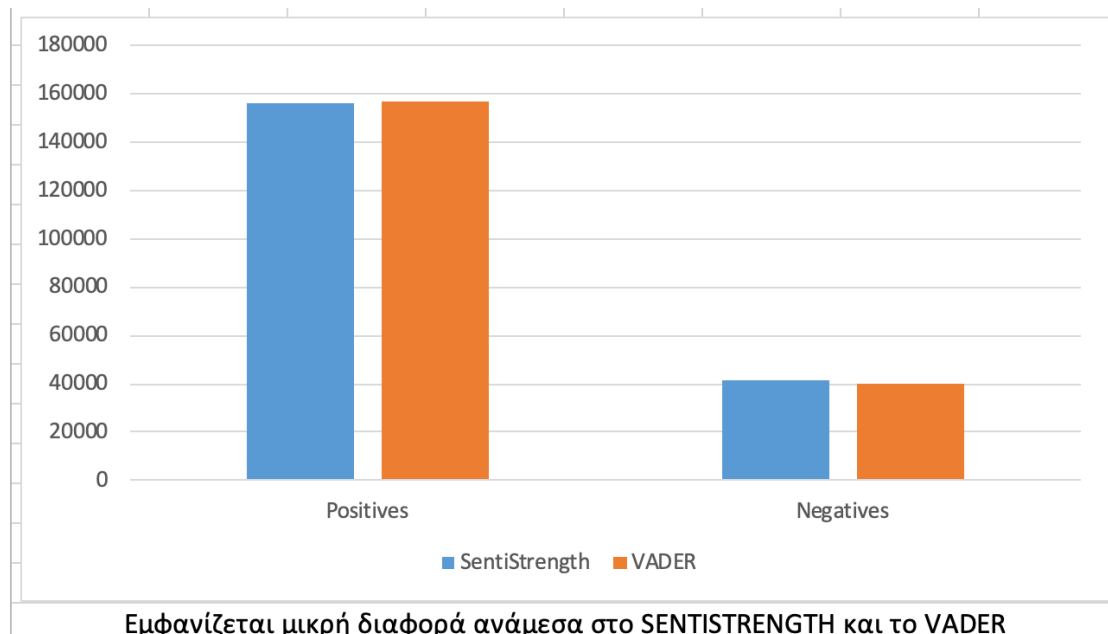
BHMA 3 SENTIMENT ANALYSIS

Χρησιμοποιήθηκαν τα Lexicon Based τρόποι εξαγωγής συναισθήματος Vader και το SentiStrength για σύγκριση και για να ερευνηθεί αν υπάρχει διαφορά στον αριθμό των θετικών και αρνητικών προτάσεων(Tweets). Στην περίπτωση του SentiStrength χρησιμοποιείται η δωρεάν έκδοση που προσφέρει το Dual Sentiment Analysis.

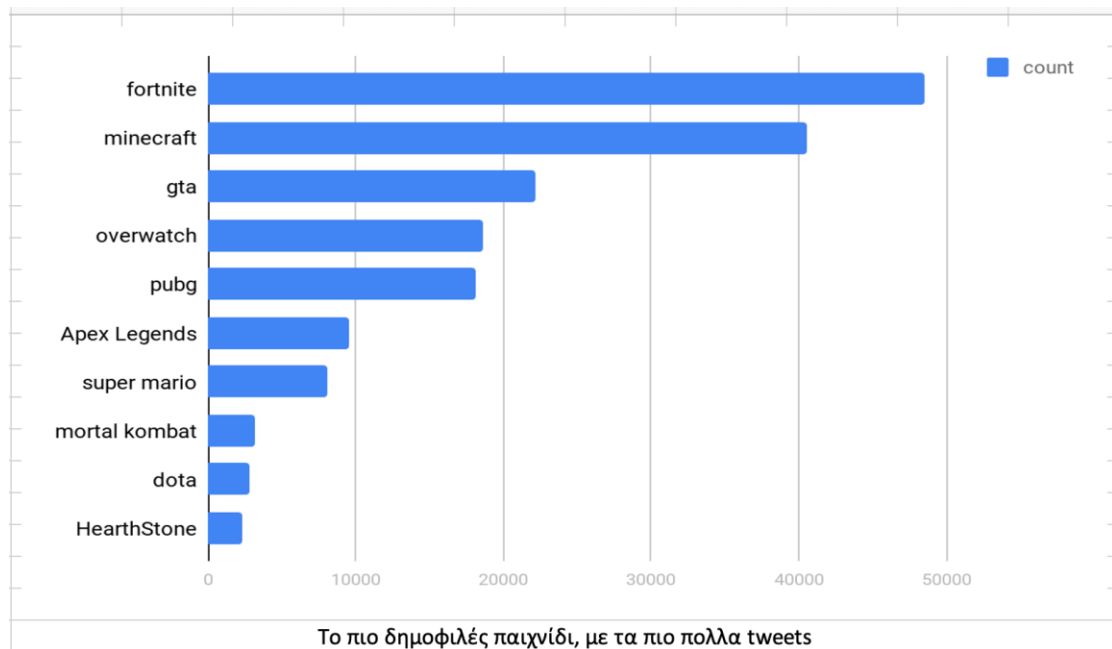
BHMA 4 DASHBOARD

Με τα δεδομένα και τα αποτελέσματα που έχουν παραχθεί από το προηγούμενο βήμα υπάρχει η δυνατότητα παρουσίασης τους. Αρχικά η σύγκριση των δύο Lexicon Based προσεγγίσεων δεν

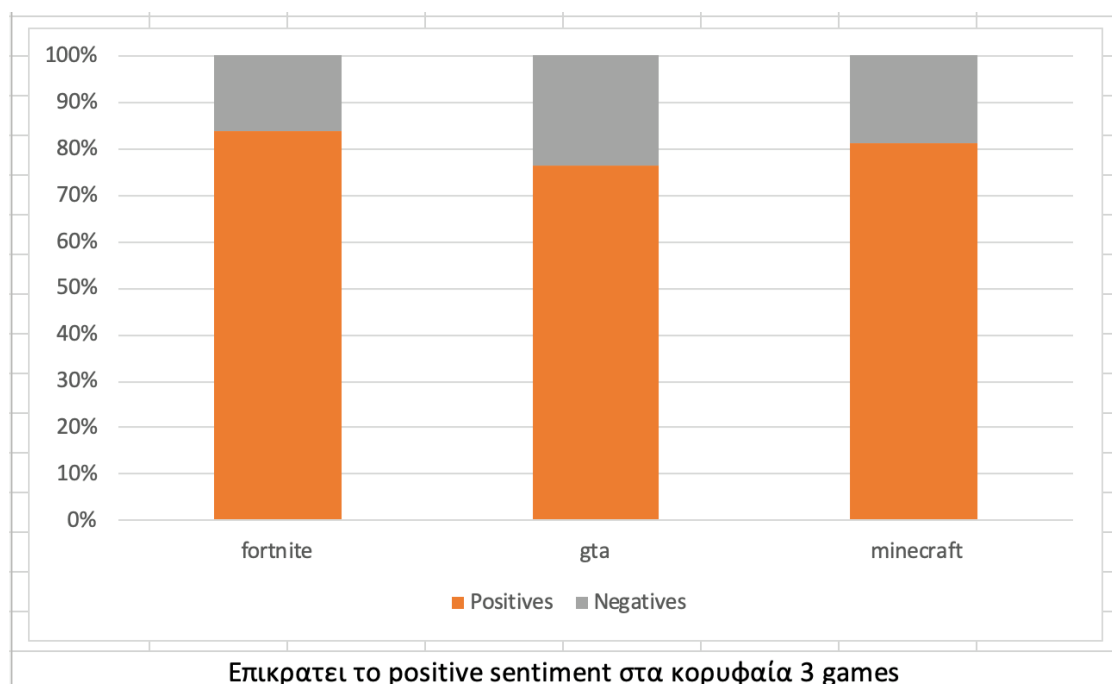
έδειξε κάποια σημαντική διαφορά όσον αφορά την εξαγωγή του Sentiment. Όπως βλέπουμε στην παρακάτω εικόνα το SentiStrength και το Vader με βάση τα υπάρχοντα δεδομένα, μας εμφανίζουν μικρή διαφορά όσον αφορά τον αριθμό των Positive και Negative προτάσεων (Tweets).



Κάποια ενδιαφέροντα συμπεράσματα που μπορούν να εξαχθούν είναι με βάση τα tweets να βρούμε ποιο είναι το πιο δημοφιλές παιχνίδι. Στην παρούσα περίπτωση το πιο δημοφιλές παιχνίδι είναι το Fortnite με διαφορά και ακολουθούν το Minecraft και το gta για να συμπληρώσουν την αρχική τριάδα.



Τέλος άλλο ένα ενδιαφέρον στατιστικό που παρουσιάζεται είναι ότι μπορούμε να διακρίνουμε τα ποσοστά Sentiment για να ερευνήσουμε ποιο είναι το επικρατέστερο. Ενδεικτικά στην παρακάτω εικόνα διακρίνεται ότι στα τρία πιο δημοφιλή παιχνίδια επικρατούν τα ποσοστά του Positive Sentiment με βάση τα tweets των χρηστών.



ΣΗΜΑΝΤΙΚΟ

Για περισσότερες πληροφορίες όσον αφορά την αναλυτικότερη επεξήγηση των τεσσάρων βημάτων, είναι ορατές στην σελίδα της ομάδας του Sentiment Analysis στο Github <https://github.com/kongeor/ihu-sentiment-analysis> .