Ανάκτηση Πληροφορίας 2022

Word Embeddings

Ομάδα

Φραγκαθούλας Χρήστος ΑΜ 4196

Λιάτσος Νικόλαος ΑΜ 4101

Το μοντέλο που χρησιμοποιήθηκε για την άσκηση είναι ένα μοντέλο της Google που είναι pre-trained



Δοκιμάσαμε κάποια παραδείγματα όπως

Σε αυτό το παράδειγμα παρατηρούμε ότι το διάνυσμα (man+computer) έχει μεγάλη ομοιότητα με το διάνυσμα (woman+computer) με ποσοστό περίπου 0,00041. Με την αντικατάσταση του τελευταίου διανύσματος σε (woman+householder) παρατηρούμε ότι η συσχέτιση μεταξύ τους είναι μεγαλύτερη σε σχέση με πριν.

Ένα επιπλέον παράδειγμα



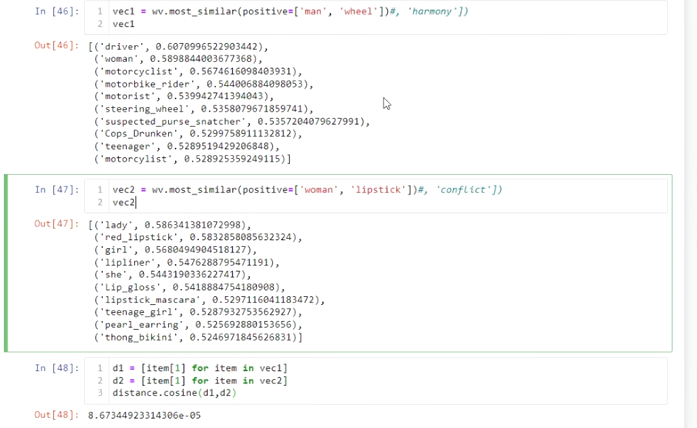
Αυτό το παράδειγμα δείχνει την προκατάληψη που υπάρχει, ότι δηλαδή οι λευκοί άντρες έχουν εισόδημα, ενώ οι μαύροι άντρες είναι ποιο εκμεταλλεύσιμοι.

Ακόμη το παρακάτω παράδειγμα

Πολλά κείμενα της google βασιζόμενα στην Wikipedia, φαίνεται ότι υπάρχει η σύνδεση της Ελλάδος με την ειρήνη και της Ρωσίας με τον πόλεμο.

Παρακάτω θα δούμε μια διαφορά μεταξύ δύο μοντέλων

Μοντέλο της Google



Άλλο μοντέλο

Τα μοντέλα πρέπει να εκπαιδεύονται με κατάλληλα κείμενα, γιατί αυτά τα μοντέλα παίζουν σημαντικό ρόλο στην καθημερινότητά μας με παράδειγμα τα παραπάνω, όπου τα 2 μοντέλα συσχετίζουν διαφορετικά τον άνδρα και την γυναίκα στην καθημερινότητάς τους. Ακόμη παρατηρούμε το ένα μοντέλο εκφράζει και συνδέει την γυναίκα με προσβλητικούς χαρακτηρισμούς, ενώ το άλλο μοντέλο τους έχει αποφύγει.

Επιπρόσθετα το επόμενο παράδειγμα



Αυτό μας δείχνει ότι η Ελλάδα έχει συσχετιστεί με το φαγητό, λόγω του πολιτισμού και η Πορτογαλία με το ποδόσφαιρο λόγω τον πολλών γνωστών ομάδων που διακατέχει.

Τέλος το παράδειγμα που φαίνεται παρακάτω

Παρουσιάζει την συσχέτιση που έχει η δημοκρατία με την ελευθερία όσο και ο Βασιλιάς με την σκλαβιά.