פרויקט גמר תומר ואור

רעיון הפרויקט הוא לנסות להבין מה הדרך הכי טובה לכסות את כל השפה בהינתן שני כלים, glove ו-word2vec.

שלב 1: להכיר את שני הכלים לעומק, למדנו להשתמש בשני הכלים בצורות שונות על מנת להביא ליותר אפשרויות (וקטורים)

שלב 2: עיבוד המידע-data set, ניסינו להוריד את תחיליות ורישות מהמילה על מנת להגיע לתמציתה, ניסיון זה לא עבד, רק 20 אחוז מתוך כל המידע יצא הגיוני. ולכן התחלנו להפעיל תותחים יותר כבדים.

ניסיון נוסף של עיבוד המידע הוא לנסות לעשות הפרדה של רישות וסיפות (לא למחוק אלא לשים רווח) ואז עליהם לעבוד.

ניסיון נוסף הוא להפריד בין אותיות ואז לעבוד אליהם.

כמובן שניקח רק את האפשרות שתיהיה הכי טובה, גם מבחינת כמות וגם מבחינת איכות.

לאחר ניסוי וטעייה זה נוכל להתקדם לשבל הגדול הבא והוא החלפת מילים בעייתיות, ושילוב של ווקטורים אחרי עיבוד

///// אני לא ימשיך אבל בעיקרון זה אומר אפשרויות של חיבור כפל וכו'