פרטים טכנים על הפרויקט:

1. נערך פיילוט קטן לבדיקת התכנות – 5 אנשים \* 100 משפטים

עיבוד מקדים:

1. בפרויקט עצמו - 290 HITS המורכבים כל אחד מ:
   1. 70 משפטים עם טעויות שתויגו רק על ידי אדם זה.
   2. 15 משפטי בקרה עם טעויות (מתוך מאגר של 200 כאלו סה"כ).
   3. 15 משפטים ללא טעויות ("מושלמים")
2. המשפטים שנבחרנו הינם:
   1. משפטים עם לפחות 7 מילים (כאשר ההגדרה של מילה פה היא כל דבר שהופיע בקורפוס בין שני רווחים)
   2. משפטים שלא מכילים את התווים או המחרוזות הבאות: ";", "\*", "[", "]", "&", "http"
3. המשפטים עברו את הניקוי הבא:
   1. מחיקת התו רווח אם הוא הגיע לפי אחד מהתוים הבאים: , . / ) ! % $
   2. מחיקת התו רווח אם הוא הגיע אחרי התו: )

עיבוד אוחר:

1. כלל הציונים נורמלו באמצעות z-score כך שציוניו של כל אדם (או HIT?) בעלי ממוצע 0 ושונות 1.
2. נפסלו הציונים שניתנו על ידי אנשים ש:
   1. הציון הממוצע שלהם על משפטים "מושלמים" גדול מ-0 באופן מובהק (כלומר גרוע מהממוצע)
   2. מילאו מהר מידי