<u>עבודה 3:</u>

מגישים:

נתנאל לוי 208463158 דור ירון 316083542 גיא ידבבסקי 205773252

Reports:

Design:

Map-reduce explanation:

Step1:

הקלט הינו הקובץ של ה biarcs מפעילים stemmer על הזוגות של שמות העצם ומוצאים את המסלול בין כול שני שמות עצם

< noun1,noun2 | route amount > הפלט ניראה כך

Step2:

מאגדים את כול המסלולים השונים עבור זוג שמות עצם

הפלט ניראה כך

< noun1,noun2 | route1 amount1 route2 amount2.... >

כעת חוזר לנו קובץ שבכול שורה שלו יש זוג מילים וכל המסלולים השונים שבהם זוג המילים האלו מופיע

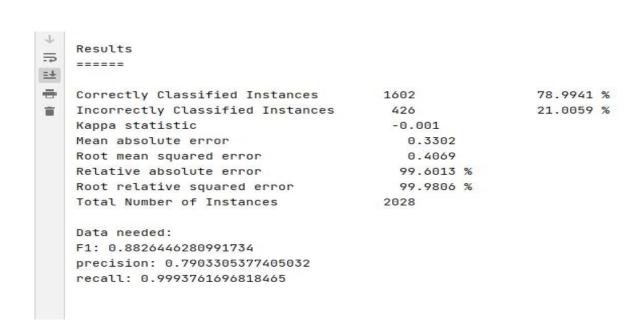
במחלקה בשם CreateVectors אנו מבצעים את הדברים הבאים:

- 1. מפעילים stemmer על הקובץ
- 2. סופרים את מספר המסלולים השונים בקובץ שחזר מה map-reduce (מספר המסלולים השונים יהיה גודל הווקטור)
 - txt איוצרים קובץ של הווקטורים בפורמט של 3.
- vectors-with- בעזרת שינוי קטן בקוד שאנחנו מכניסים ניתן לקבל גם קובץ -4 nouns.txt

לאחר מכן כאשר יש לנו ביד את הקובץ vectors.txt אנו ממירים אותו לקובץ csv. בעזרת excel ואחר מכן ממירים אותו לקובץ arff בעזרת

המחלקה בשם Weka אנו מבצעים את הדברים הבאים:

- 1. מפעילים את ה DecisionStump) classifier) בעזרת הספרייה של
 - F1 ,Recall and Precision מוצאים את הערכים של.
- TruePositive ,TrueNegative ,FalsePositive ,FalseNegative ,4 מוצאים 5 ערכים של



Communication:

מצורף קובץ syslog עם מידע לגבי מספר ה syslog

True Positive: 0

True Negative: 1602

book, navig

cent,sand

conserv, activ

enterpris, organ

fact,room

Iondon,citi

nation,union

offic,work

False Positive: 1

hope,expect

False Negative: 425

author, author

christian,belief

east,asia

general, aspect

kind, director

mission,work

rem,sleep

sphere,boundari

woodland, forest