

עבודה 3:

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 ומוציאים את המסלול בין כול שני שמות עצם

הפלט נראה כך $\langle \text{noun1}, \text{noun2} \mid \text{route amount} \rangle$

Step2:

מאגדים את כול המסלולים השונים עבור זוג שמות עצם

הפלט נראה כך

$\langle \text{noun1}, \text{noun2} \mid \text{route1 amount1 route2 amount2} \dots \rangle$

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

2359	abandon,instanc	X/NN/attr/IN/prep/NN/pobj/Y	172
2360	abandon,institut	X/NN/dobj/IN/prep/NN/pobj/Y	41
2361	abandon,insur	X/NN/nsbj/VBN/rcmod/NN/nsbjpass/Y	78
2362	abandon,intent	X/NN/dobj/IN/prep/NN/pobj/Y	47
2363	abandon,interv	X/NN/pobj/IN/prep/NN/pobj/Y	62
2364	abandon,invas	X/NN/pobj/IN/prep/NN/pobj/Y	33
2365	abandon,issu	X/NN/attr/IN/prep/NNS/pobj/Y	64
2366	abandon,job	X/NN/ROOT/IN/prep/NNS/pobj/Y	32
2367	abandon,joy	X/NN/dep/TO/prep/NNS/pobj/Y	39
2368	abandon,junctur	X/NN/pobj/IN/prep/NN/pobj/Y	38
2369	abandon,land	X/NN/pobj/IN/prep/NNS/pobj/Y 74---X/NN/nsbj/IN/prep/NN/pobj/Y 71---X/NN/dobj/IN/prep/NN/pobj/Y 82---X/NN/conj/IN/prep/NN/pobj/Y 113---X/NN/pobj/IN/prep/NN/pobj/Y 409	
2370	abandon,langvag	X/NN/dobj/IN/prep/NN/pobj/Y	79
2371	abandon,laughter	X/NN/conj/IN/prep/NN/pobj/Y 31---X/NN/pobj/IN/prep/VBG/pcomp/NN/dobj/Y	42

במחלקה בשם CreateVectors אנו מבצעים את הדברים הבאים:

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

המחלקה בשם Weka אנו מבצעים את הדברים הבאים:

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

↓	Results		
↺	=====		
≡↓			
🖨	Correctly Classified Instances	1602	78.9941 %
🗑	Incorrectly Classified Instances	426	21.0059 %
	Kappa statistic	-0.001	
	Mean absolute error	0.3302	
	Root mean squared error	0.4069	
	Relative absolute error	99.6013 %	
	Root relative squared error	99.9806 %	
	Total Number of Instances	2028	
	Data needed:		
	F1: 0.8826446280991734		
	precision: 0.7903305377405032		
	recall: 0.9993761696818465		

Communication:

מצורף קובץ syslog עם מידע לגבי מספר ה Pairs שנשלחו

True Positive: 0

True Negative: 1602

book,navig

cent,sand

conserv,activ

enterpris,organ

fact,room

london,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