

מבוא לעיבוד שפה טבעית

מטלה 12 – פתרון

אייל טרבלסי, ת.ז. 200953859

רומן ציפרון, ת.ז. 306591645

שאלה 1

סעיף	פרמטר	קובץ אימון	קובץ גולד
1	מופעי יוניגרם של סגמנטים	127884	11282
2	סוגי יוניגרם של סגמנטים	15986	3171
3	מופעי סגמנט-תג (זזה ל-1 כי לכל סגמנט יש תג)	127884	11282
4	סוגי סגמנט-תג	37	36

5. מדד העמימות עבור קובץ האימון: 1.13493056424

מדד העמימות עבור קובץ הגולד: 1.07978555661

6.