



תרגיל מעבדה 5

בהמשך לתרגיל המעבדה, ממשו את השיפורים הבאים האפשריים:

1. ייצוג tf

השתמשו בנוסחה: $wf_{i,d} = 0$ if $tf_{i,d} = 0$, $1 + \log tf_{i,d}$ otherwise

2. ייצוג tf-idf

השתמשו בנוסחה: $w_{i,d} = tf_{i,d} \times \log(n / df_i)$

3. שימוש בדמיון קוסינוס במקום במרחק אוקלידי (predict) (בפונקציית predict)

• השתמשו ב \log בבסיס 10

בכל שלב דווחו:

1. את השיפור בדיוק בהשוואה ל baseline (שימוש בייצוג Boolean עם מרחק אוקלידי).
2. את השיפור בדיוק בהשוואה לשלב הקודם. לדוגמה: מהו השיפור שהושג (אם בכלל) בשינוי של כל המילים ל lowercase והורדת סימני פיסוק בהשוואה לשלב הקודם – שימוש בייצוג tf-idf.
3. הסבירו את הסיבה לשינויים בדיוק של האלגוריתם בכל אחד מהשלבים.

הדפיסו את הפלט למסך באופן הבא:

Accuracy results:

Boolean: 0.000

tf: 0.000

tfidf: 0.000

tfidf with cosine similarity: 0.000



דגשים נוספים :

- (1) ניתן להשתמש בקוד שכבר מימשתם בתרגילי הבית.
- (2) ניתן להוסיף פונקציות נוספות, במידה ותמצאו לנכון.
- (3) עליכם לכתוב את הקוד בהתאם לדרישות של PEP 8. יותר על זה [כאן](#) או [כאן](#). אתם לא חייבים לעמוד בכל הדרישות של PEP, אבל יש כמה חשובות לנו. יותר על זה במסמך במודל.

- אורך שורה לא יותר מ-120
- שמות של משתנים/טיפוסים/קבצים
- Imports
- Blank lines

(4) יש לתת שמות בעלי משמעות לכל משתנה.

(5) חובה לתעד את הקוד באנגלית.

הוראות הגשה

- התרגיל להגשה בזוגות או ביחידים (נדרש אישור להגשה ביחידים).
- לפני ההגשה, חובה לוודא שהתוכנית עובדת במעבדת ההוראה ולא בסביבה אחרת.
- ההגשה חייבת להכיל קובץ אחד (קובץ zip) :
 1. שם הקובץ חייב להיות hw5_lab_xxxxxxxx_yyyyyyyy.zip כאשר xxxxxxxx ו-yyyyyyyy הם מספרי תעודות הזהות של המגישים, כולל ספרת ביקורת.
 - הקובץ מכיל את כל קבצי הקוד וקובץ דו"ח שלכם עם תשובות לשאלות. אין להכיל תיקייה ובתוכה קבצי הקוד, אלא את קבצי הקוד עצמם.
- **הערה :** עליכם לוודא שהתוכנית מתחילה לפעול מקובץ "main.py" בלבד.
- ההגשה היא אלקטרונית בלבד, דרך אתר ה-moodle של הקורס. תרגילים שיוגשו בכל דרך אחרת לא ייבדקו.
- אין להגיש את אותו הקובץ פעמיים. התרגיל יוגש ע"י אחד מבני הזוג.
- שימו לב שההגשה תיחסם בדיוק בשעה 23:55 ביום ההגשה. מומלץ להגיש לפחות שעה לפני המועד האחרון.
- ניתן להגיש כמה פעמים. רק ההגשה האחרונה תישמר.
- תרגיל בית שלא יוגש לפי הוראות ההגשה – לא ייבדק (כלומר יקבל ציון 0).
- לצורך תרגיל הבית יפתח פורום. ניהול שאלות ומתן תשובות בנושא התרגיל יתבצע דרך הפורום בלבד.

בהצלחה!