**קובץ תיעוד לפרוייקט 2 - ניהול נתונים באינטרנט**

לירון כהן 207481268, יובל מור 209011543

**תיאור הקוד**

חלק א' - בניית ()

* עבור כל קובץ בתיקייה הנתונה:
  + פרסור הקובץ ל-
  + מציאת כל המסמכים () בקובץ
  + עבור כל מסמך:
    - חילוץ שדה ה-
    - חילוץ כל המילים במסמך ()
      * השדות שבחרנו לחלץ והובילו לתוצאות טובות הם:
    - עבור כל מילה במסמך:
      * ביצוע , הסרת סימני פיסוק, ניפוי והמרה ל-.
      * שמירת המילה במילון המילים ().
      * העלאת ה- של המילה במילון המסמכים תחת המסמך המתאים.
      * עדכון ערך ה- של המסמך בעת הצורך.
    - חישוב ערכי ה- של המילים במסמך ():
      * עבור כל מילה במילון המילים, אם היא נמצאת במסמך, נחלק את הערך שלה ב- של המסמך.
* חישוב ערכי ה- של המילים ():
  + חישוב , שהוא כמות המסמכים.
  + עבור כל מילה במילון המילים נבצע את חישוב הנוסחה לקבלת ה- מתוך ה- ו-.
* חישוב ערכי המשקלים של המילים ():
  + הכפלת ה- ב- שחושב לעיל לקבלת ערך ה- של כל מילה.
* שמירת מילון המסמכים ומילון המילים לקובץ ().

חלק ב' - אחזור מידע בהינתן שאלה ()

* טעינת מילון המסמכים ומילון המילים מקובץ ה- ().
* פרסור השאלה ()
* אם מדובר ב-, חישוב ציונים בהתאם ():
  + עבור כל מסמך:
    - חישוב ערכי באמצעות הנוסחה המתאימה ().
    - אם התוצאה מעל מסוים (הקבוע שנבחר לאחר אופטימיזציות להיות 0.08), נכניס את התוצאה למילון .
  + מיון המילון בסדר יורד לפי הערכים שהתקבלו והחזרתו.
* אם מדובר ב-, חישוב ציונים בהתאם ():
  + חישוב הקבועים הכלליים הנדרשים לנוסחה ().
  + עבור כל מסמך:
    - חישוב ערך ה- המתאים.
    - חישוב ערך ה- באמצעות הנוסחה המתאימה ().
    - אם התוצאה מעל מסוים (הקבוע שנבחר לאחר אופטימיזציות להיות 0.08), נכניס את התוצאה למילון .
  + מיון המילון בסדר יורד לפי הערכים שהתקבלו והחזרתו.
* שמירת התוצאות לקובץ טקסט ().