תיק למתכנת: **Tiger Team**

המסמך הזה מחולק ל-2 חלקים, החלק ראשון הוא חלק ה-Backend, שמסביר על המחלקות ועל המיקרו שירותים שהשתמשנו בהם בזחלן ובהעלאה של התוכן לשרת Firebase. החלק השני הוא חלק ה-Frontend, שמסביר עלהצגת המידע הדפים למשתמש בניהם: התחברות, חיפוש עדכון ומחיקת נתונים ע"י משתמש מסוג אדמין, צ'אט AI, סטטיסטיקות מעניינות.

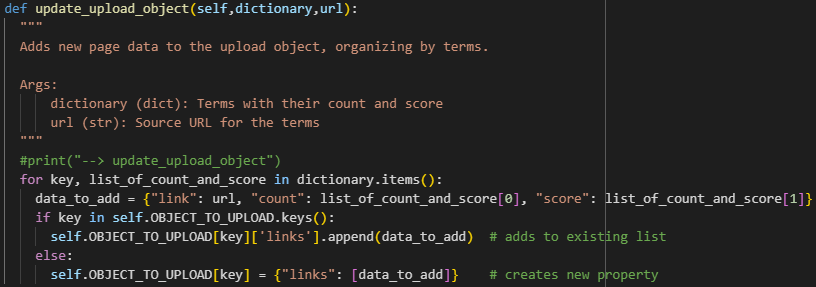
חלק 1 - **Backend:**

1. מחלקת **UploadObjectService**

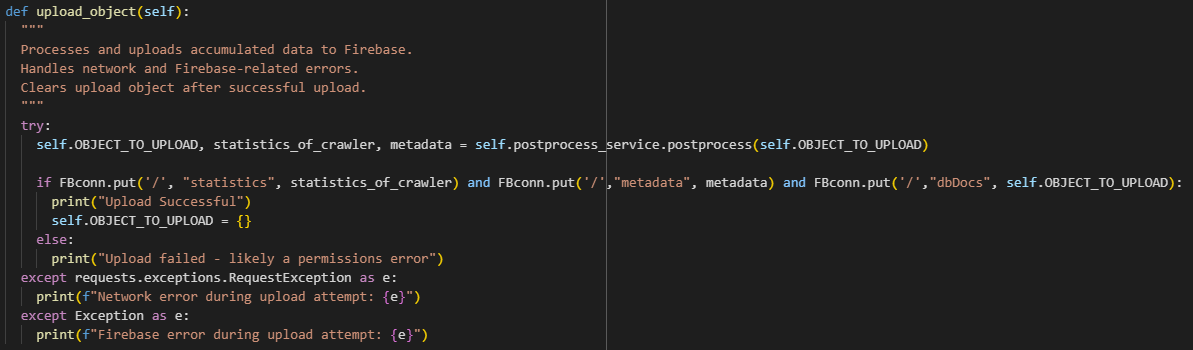


מיקרו שירות שמאפשר לעדכן אובייקט ששמור במחלקה ובסוף להעלות אותו לשרת Firebase.

פונקציה **Update\_upload\_object:**הפונקציה מקבלת לינק לדף ספציפי ואת המילון שלה ומעדכנת את האובייקט OBJECT\_TO\_UPLOAD, היא בעצם מכניסה לכל מילה במילון את הלינק הנוכחי, את כמות המופעים של המילה בדף, ואת ה-score שהיא קיבלה בדף זה.



פונקציה **upload\_object:**הפונקציה מעלה בפעימה אחת את האובייקט OBJECT\_TO\_UPLOAD לשרת Firebase. במידה והיא הצליחה היא מאפסת את האובייקט.



1. **מחלקת DictionaryCreationService**



מיקרו שירות שמאפשר למשתמש ליצור מילון לכל המילים שקיימות בדף.

פונקציה **create\_count\_dictionary:**  
יוצרת מילון לכל מילה בדף. המילון מאותחל בצורה הבאה:

{  
“word”: [count, [index1, index2, …, indexN]  
 “anotherWord”: [count, [index1, index2, index3]  
}

כאשר count מסמל את כמות המופעים של המילה בדף, וכל index מסמל את מספר המילה בדף.

A screen shot of a computer

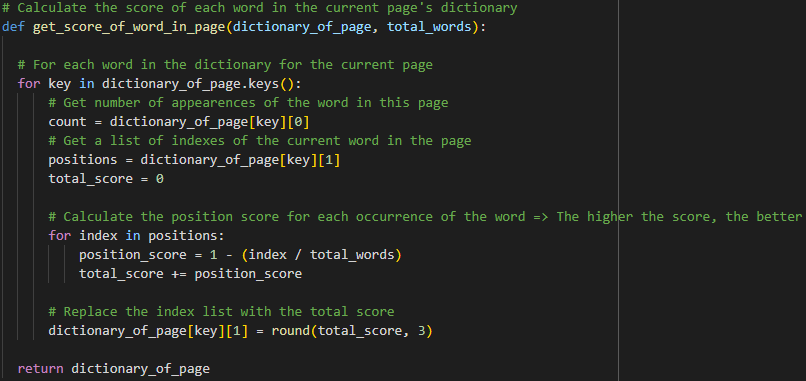
Description automatically generated

פונקציה **get\_score\_of\_word\_in\_page:**פונקציה לחישוב תיעדוף אינדקס. האלגוריתם מחשב ציון של מילה בדף כתלות במיקום שלה. הוא מתבצע בסדר הבא:

* עבור כל מילה בדף, חשב את הציון שלה לפי הנוסחה:

כאשר מסמל את המיקום של המילה בדף מתוך כל המילים בדף, ו- מסמל את מספר המילים הכולל בדף הנוכחי.

* סכום את כל המופעים של אותה המילה במשתנה .
* עדכן במילון של הדף את הערך למילה הנוכחית.

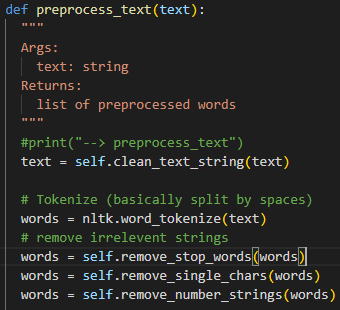


1. **מחלקתPreprocessService**



מיקרו שירות שמחזיק פונקציות לניקוי, ועיבוד טקסט.

**:preprocess\_text**פונקציה



פונקציה ראשית, מריצה את כל שאר הפונקציות לעיבוד הטקסט. היא מבצעת ניקוי של מילים בודדות, מספרים, מילות קישור, סימנים כגון ",.[)" ועוד.

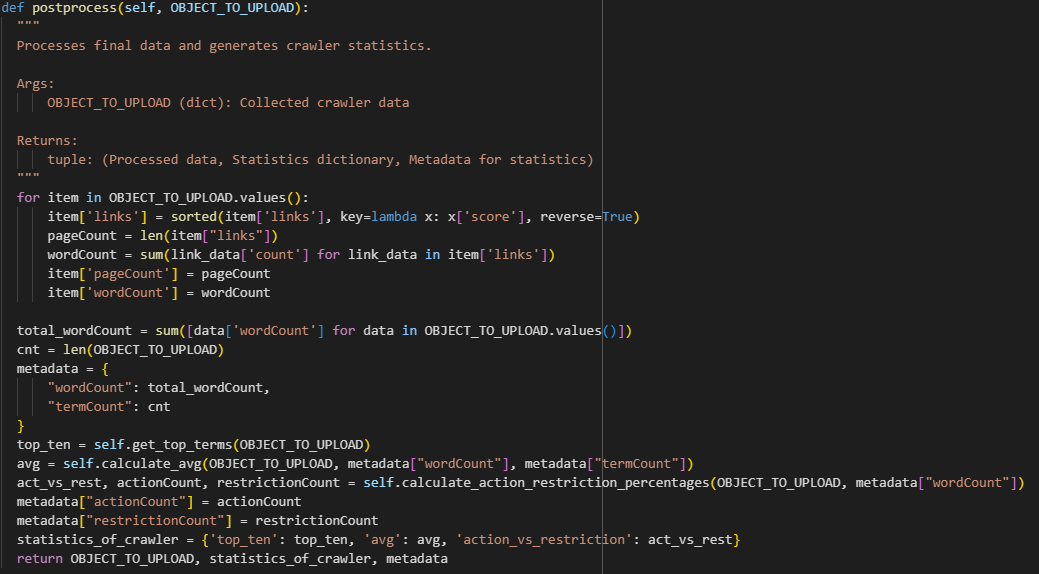
1. **מחלקת PostprocessService**

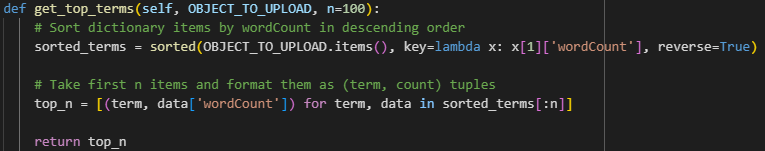


מיקרו שירות שמחזיק פונקציות לניקוי, ועיבוד טקסט.

**פונקציה postprocess:**

* פונקציה ראשית, מריצה את כל שאר הפונקציות לעיבוד הטקסט. היא מבצעת ניקוי של מילים בודדות, מספרים, מילות קישור, סימנים כגון ",.[)" ועוד.
* אחראית על בניית מידע עבור הצגת הסטטיסטיקות המעניינות ועריכת הסטטיסטיקות המעניינות.

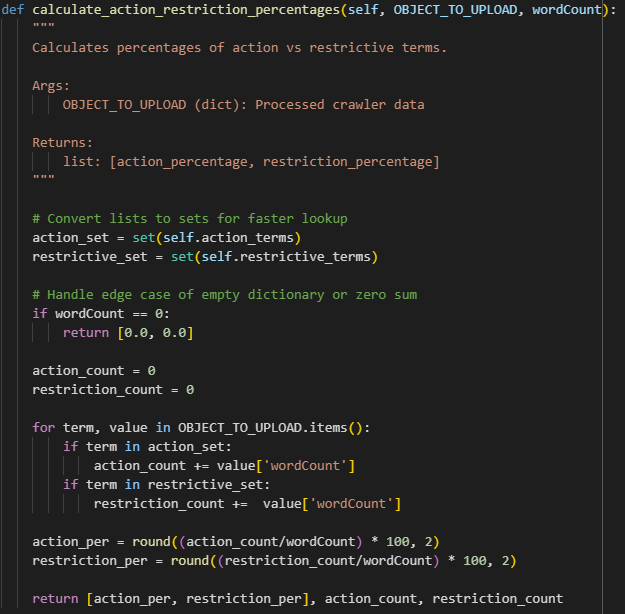


**פונקציה get\_top\_terms:**הפונקציה ממיינת את המילים לפי כמות המופעים שלהן בכל האתר ומחזירה רשימה של n זוגות מילים וכמות המופעים שלהן בכל האתר (n נתון לבחירה בקריאה לפונקציה, בברירת המחדל n מוגדר להיות 100(.

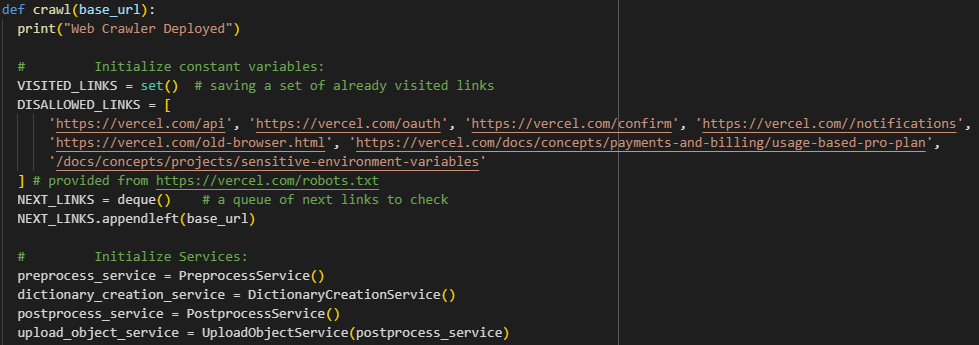
פונקציה**calculate\_avg :**מחשבת כמה פעמים בממוצע מונח כללי מופיע בדף.



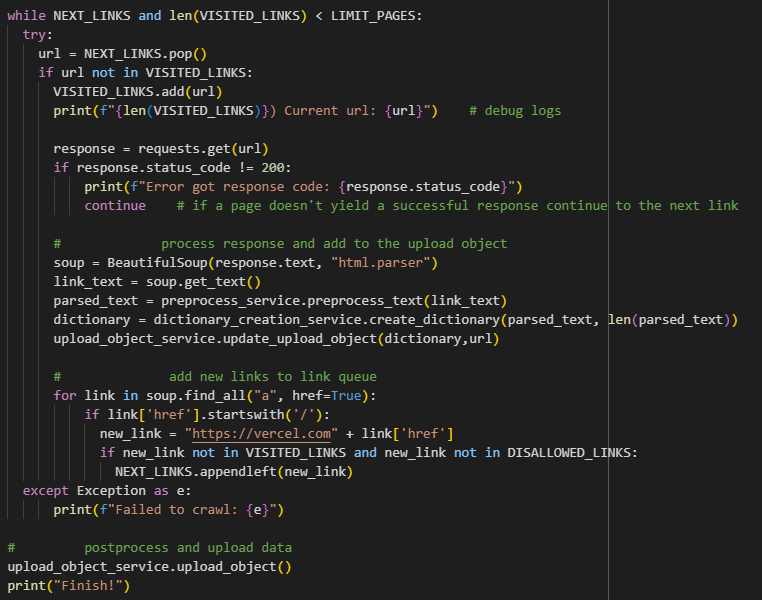
פונקציה **calculate\_action\_restriction\_percentages:**הפונקציה מחשבת מהו האחוז מכלל המילים שהם מילות action, לעומת מילות restriction. היא בודקת אם מילה נמצאת בסט action\_set או restrictive\_set שמוגדרים מראש במחלקה, ומחזירה את אחוז המופעים של מילים אלו, ומספר המילים האלו.



הזחלן: פונקצית **crawl**פונקצית הזחלן מקבלת את הלינק של דף הבית שממנו אפשר להתחיל את החיפוש. בזחלן מוגדרים הלינקים אליהם אסור לזחלן להכנס, והלינקים האלו נלקחו מהקובץ robots.txt, כדי למנוע חסימה של הזחלן באתר של Vercel. הפונקציה מגדירה את ארבעת המחלקות השונות שעליהם הוסבר מעל.



לאחר מכן, בפונקציה מוגדרת לולאת while אשר עוברת על כל הלינקים שהיא מוצאת בדף הנוכחי, ומוסיפה אותם למחסנית שאליה הוא ניגש כדי לדעת באיזה דף לבצע זחילה. הפונקציה שומרת את הלינקים שבהם כבר ביקרנו כדי למנוע לולאה אין סופית של זחילה. ישנו משתנה גלובלי LIMIT\_PAGES שמוגדר ל-300 דפים באתר.



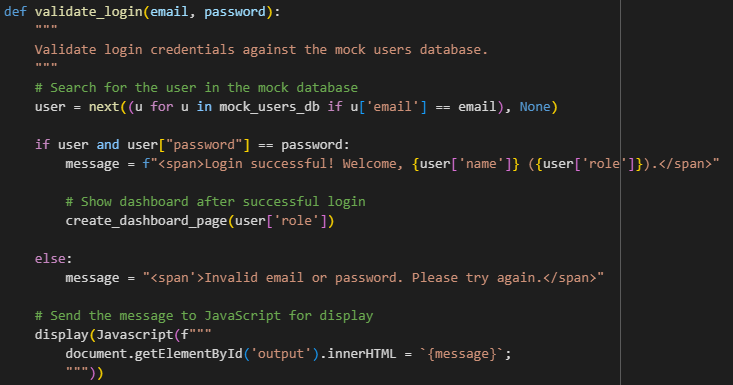
חלק 2 – **Frontend:**

**1. LoginPage**

מאפשר למשתמש להתחבר למערכת באמצעות אימות פרטים מול מסד נתונים מדומה.

**פונקציית validate\_login:**

הפונקציה מקבלת אימייל וסיסמה ומאמתת אותם מול רשימת משתמשים מוגדרת מראש. היא מחזירה הודעת הצלחה או כישלון בהתאם לתוצאת האימות. במקרה של הצלחה, היא מעבירה את המשתמש לדף הדשבורד עם ההרשאות המתאימות לתפקידו (מנהל או לקוח).



**2. DashboardPage**

מספק ממשק ניהול ראשי המציג את האפשרויות העומדות בפני המשתמש.

הדף מכיל שלושה כפתורים עיקריים:

**Logout**: יציאה מהמערכת וחזרה לדף ההתחברות.

**Search**: מעבר לדף החיפוש.

**Statistics**: מעבר לדף הסטטיסטיקות.



1. **SearchPage**

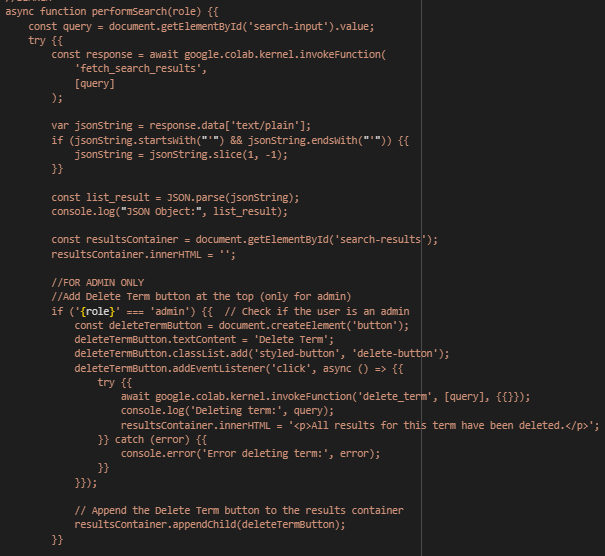
מאפשר למשתמש לחפש מילים באתר Vercel ולקבל את תוצאות החיפוש.

**פונקציית performSearch:**

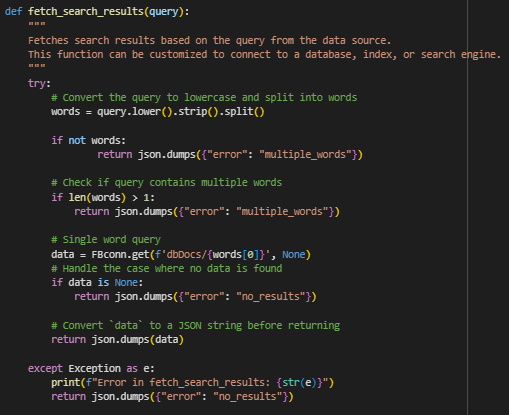
- מבצעת חיפוש של מילה בודדת במסד הנתונים

- מציגה את התוצאות בטבלה הכוללת: קישור לדף, מספר המופעים, וציון הרלוונטיות

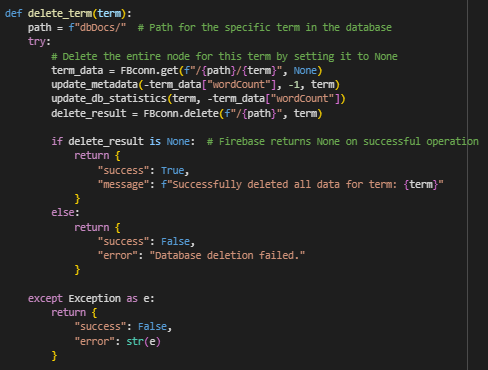
- למשתמש מנהל מוצגים כפתורי עריכה ומחיקה לכל תוצאה



משיכת תוצאות השאילתא מה-DB:



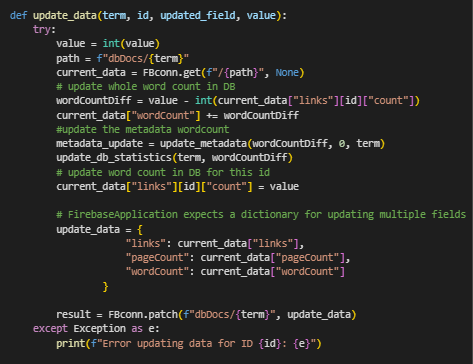
מחיקת מונח שלם מה- DB וקריאה לעדכון סטטיסטיקות בהתאם:



מחיקת מופע של מונח מה-DB וקריאה לעדכון סטטיסטיקות בהתאם:

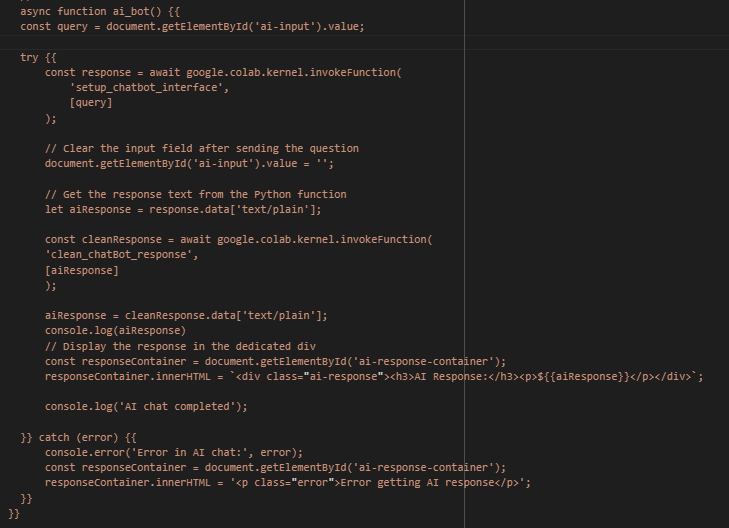


עדכון מספר המופעים של מונח בלינק וקריאה לעדכון סטטיסטיקות בהתאם:



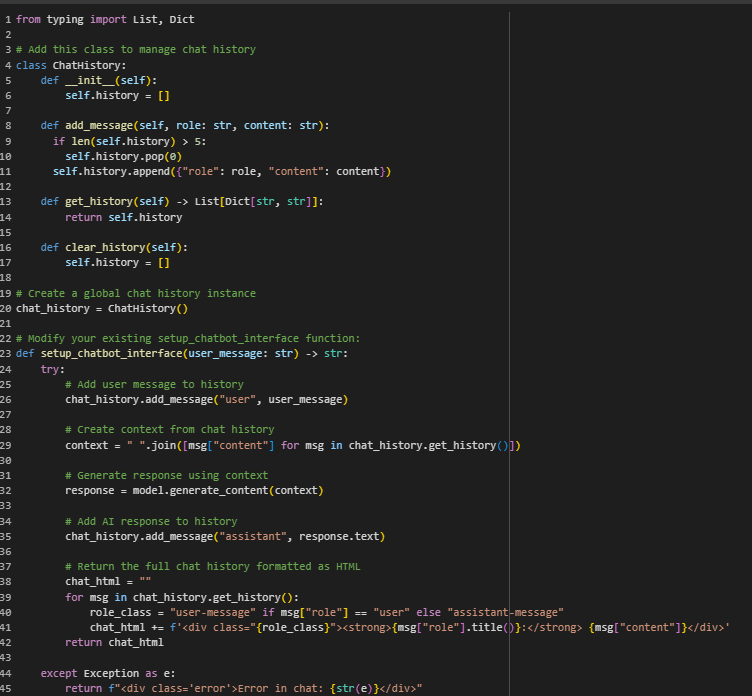
**פונקציית** **ai\_bot:**

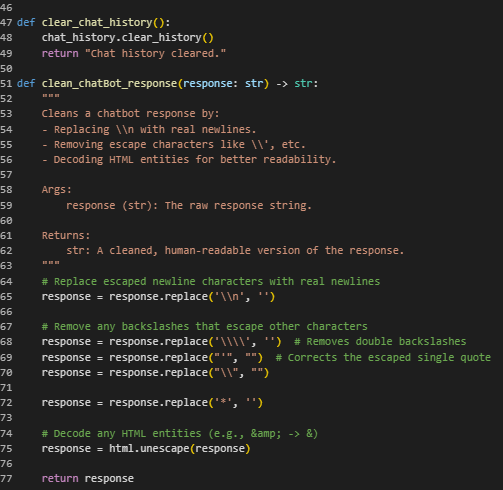
מספקת שיחה עם המשתמש על כל נושא שהמשתמש בוחר לדבר איתו באמצעות מודל Gemini של Google.



**מחלקת ChatHistory:**

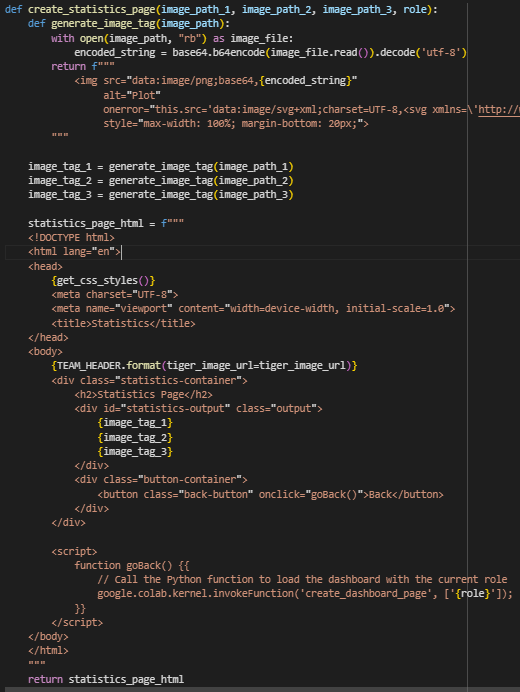
אחראית על שמירת ההיסטוריה של השיחות עם הצ'אט בוט כדי לסמלץ שיחה אמיתית.



****

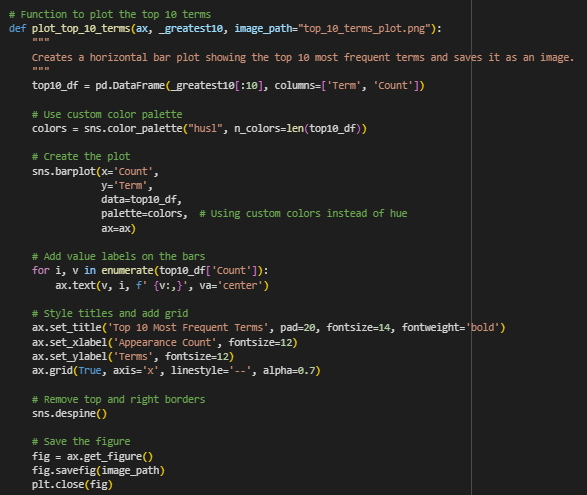
1. **StatisticsPage מחלקת**

מציג ויזואליזציות של נתונים סטטיסטיים על המילים שנאספו.

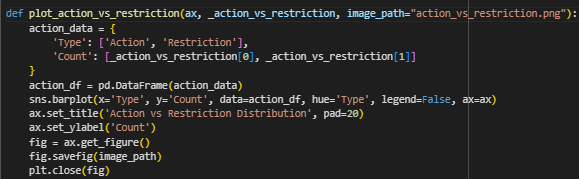


**מציג שלושה גרפים:**

**Top 10 Terms**: מציג את עשר המילים הנפוצות ביותר



**Action vs Restriction**: השוואה בין מילות פעולה למילות הגבלה



**Average Term Appearance**: ממוצע הופעות של מילים באתר.



כל הגרפים נוצרים באמצעות ספריית Matplotlib ומוצגים כתמונות PNG בדף.

1. **CSS**עיצוב ו-

המערכת משתמשת בפונקצית CSS מרכזית המגדירה:

- עיצוב בסיסי ועקבי לכל הדפים

- התאמה למסכים בגדלים שונים (Responsive Design)

- סכמת צבעים אחידה

- עיצוב כפתורים לפי סוג הפעולה (מחיקה, עדכון, חזרה וכו')