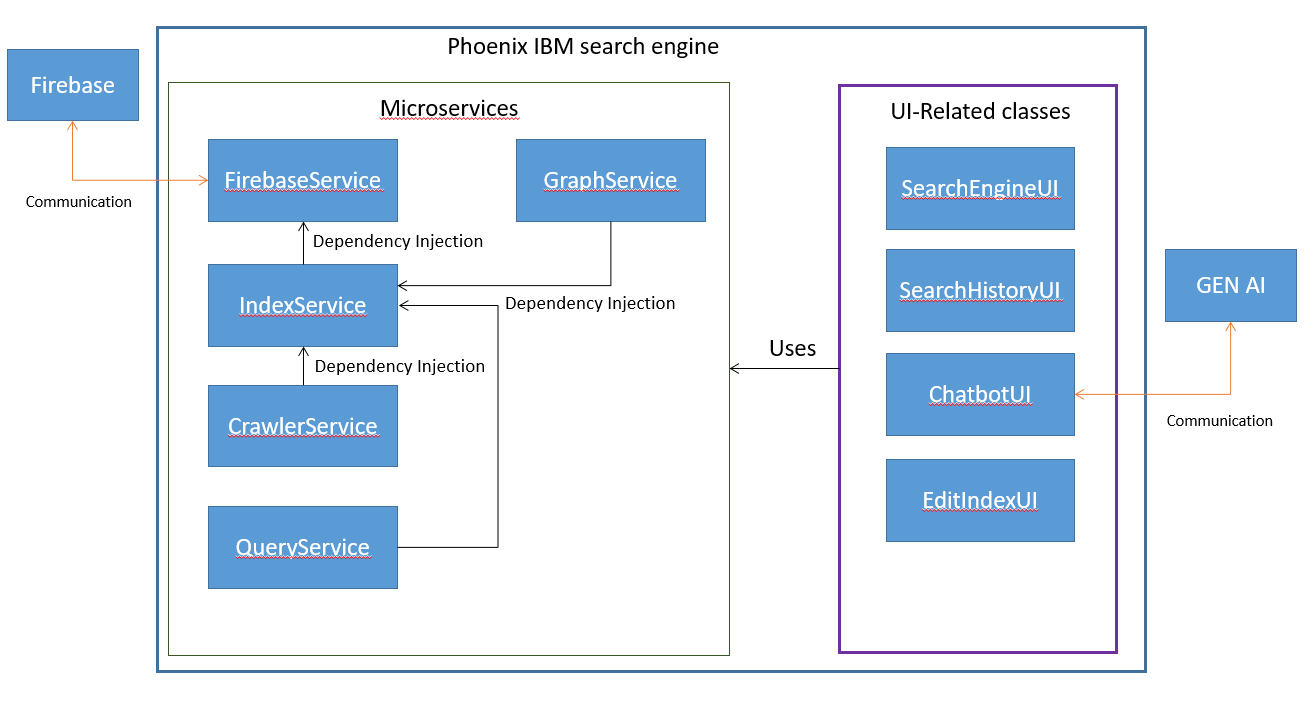
**צוות פניקס**

**תיק למתכנת:**

הקובץ המרכזי הוא הקוד בגוגל colab.

* [קישור לקובץ ב-colab.](https://drive.google.com/file/d/1jr0vu-Wt5ezfiHlkt-T8EDpim2iGSxX7/view?usp=sharing)

ארכיטקטורת המערכת:



הסבר כללי על השירותים:

1. FirebaseService – השירות שדרכו מתבצעת כל התקשורת עם ה-Firebase.

היתרון בשירות זה הוא שבגלל שכל התקשורת מתבצעת בשירות איחד, יותר קל לזהות ופתור באגים שקשורים ל-DB, ואם יש צורך בשליחת\שליפת נתונים כל מה שצריך הוא לקרוא לשירות.

1. IndexService – השירות שמנהל את האינדקס במערכת, את המילים, הקישורים שלהן, וכמות המופעים בכל קישור.

היתרון בשירות זה הוא שכל הלוגיקה של ניהול האינדקס נמצאת במקום אחד, ואם פונקציה מסוימת רוצה לבצע עדכון כלשהו או לגשת לאינדקס העדכני אז ניתן לעשות זאת דרך השירות.

1. CrawlerService – השירות שמבצע Crawling לפי עמוד בודד או על "עץ" של כתובת.

היתרון של שירות זה הוא שהשירות עושה לוגיקה נוספת "מאחורי הקלעים" כגון בדיקת ה-robot, לקיחת כתובות נוספות מהעמוד וכו'.

1. QueryService – השירות שאחראי על מתן תשובה לשאילתות, ודירוג עמודים. היתרון של שירות זה הוא שהוא עובד בנפרד משאר השירותים, הוא עדיין צריך שירות של index אבל הוא מקבל אותו על ידי dependency injection מה שמאפשר פיתוח מודולרי עם תלויות חלשות יותר.

היתרון של שירות זה הוא שלאחר האתחול על מנת לקבל תשובה צריך לשלוח רק את ה-query מבלי לבצע קריאות אחרות.

1. GraphService – השירות שאחראי על בניית הגרפים במערכת (heatmap ו-bar), הוא מקבל dependency injection של שירות אינדקס ולפיו בונה את הגרפים ומציג אותם ב-output שהוגדר לו באתחול.

היתרון של שירות זה הוא שאם רוצים לעדכן את הגרפים כל מה שצריך הוא לקרוא לו שוב, והוא יבצע את העדכון ואוטומטית יציג אותם במסכים.

* המשך בעמוד הבא.

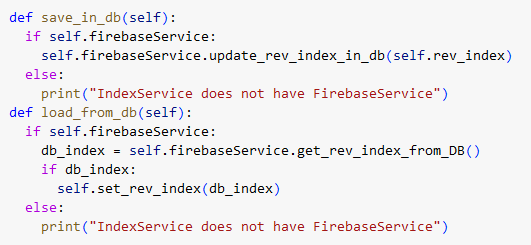
מבנה הקוד:

התקנת ספריות, ייבוא ספריות , מחלקות, ופונקציות

API של מחלקות ופונקציות מרכזיות:

|  |  |  |
| --- | --- | --- |
| **FirebaseService** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | db\_url: str = 'default\_db\_url' | אתחול המחלקה עם כתובת URL של Firebase. |
| get\_rev\_index\_from\_DB | אין | שליפת האינדקס ההפוך ממסד הנתונים. |
| update\_rev\_index\_in\_db | index\_p: dict | עדכון האינדקס ההפוך ב-Firebase. |

דוגמת שימוש:

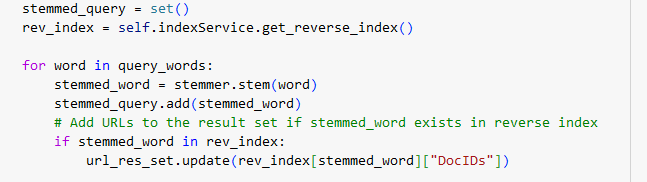


* המשך בעמוד הבא.

|  |  |  |
| --- | --- | --- |
| **IndexService** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | index: dict = None, firebaseService: FirebaseService | אתחול אינדקס ושירות Firebase. |
| process\_soup | url: str, soup: BeautifulSoup | עיבוד תוכן HTML כדי להוסיף מונחים לאינדקס. |
| get\_reverse\_index | אין | מחזיר את האינדקס ההפוך. |
| get\_url\_index | url: str | מחזיר מידע באינדקס עבור URL מסוים. |
| set\_rev\_index | newRevIndex: dict | טוען אינדקס חדש. |
| add\_new\_word | word: str | מוסיף מילה חדשה לאינדקס. |
| remove\_word | word: str | מסיר מילה קיימת מהאינדקס. |
| remove\_url | url: str | מסיר URL מהאינדקס. |
| save\_in\_db | אין | שומר את האינדקס ב-Firebase. |
| load\_from\_db | אין | טוען אינדקס מ-Firebase. |

דוגמת שימוש:

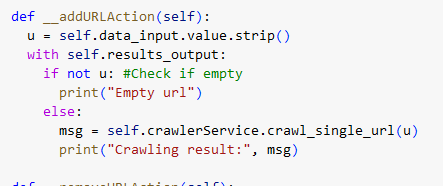


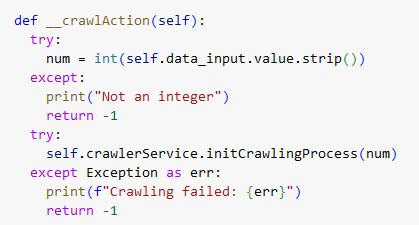


* המשך בעמוד הבא.

|  |  |  |
| --- | --- | --- |
| **CrawlerService** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | indexService: IndexService, baseURLs: set,  maxDepth: int = 100 | אתחול המחלקה עם הגדרות לזחילה. |
| initCrawlingProcess | max\_depth: int = 10 | מתחיל תהליך זחילה. |
| crawl\_website | base\_url: str, max\_pages: int = 100 | זוחל אתר בודד ומחזיר רשימת כתובות. |
| crawl\_single\_url | url: str | זוחל כתובת URL בודדת. ומחזיר הודעה עם התוצאה |

דוגמת שימוש:

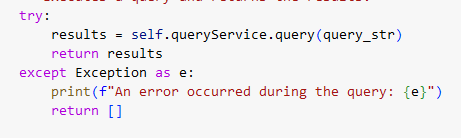




* המשך בעמוד הבא.

|  |  |  |
| --- | --- | --- |
| **QueryService** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | indexService: IndexService | אתחול המחלקה עם שירות אינדקס. |
| query | query: str | מחפש ומחזיר תוצאות ממוינות לפי ה-rank בסדר יורד, התוצאות בפורמט: |
| get\_history | אין | מחזיר היסטוריית חיפושים. |
| rank\_url | url: str, query\_words: set | מדרג כתובת URL על סמך מילות חיפוש. |
| get\_snippet | url\_val: str | מחזיר snippet עבור כתובת URL. |

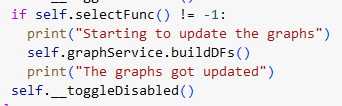
דוגמת שימוש:



|  |  |  |
| --- | --- | --- |
| **GraphService** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | indexService: IndexService, heatmap\_output: widgets.Output, bar\_output: widgets.Output | אתחול המחלקה ליצירת גרפים. |
| buildDFs | אין | בונה את מסגרות הנתונים לניתוח ומציג אותם מחדש ב-output שהתקבל באתחול. |
| build\_heatmap | אין | מחזיר גרף חום של מונחים משותפים. |
| build\_barChart | אין | מחזיר גרף עמודות של שכיחות מונחים. |

* הנתונים שמוצגים בגרפים עוברים סינון לפי כמות מנמלית של url-ים משותפים \ כמות מופעים באתר.

דוגמת שימוש:



* המשך בעמוד הבא.

מחלקות ה-UI:

מחלקות אלו מקבלות קלט מהמסכים ובעזרת השירותים מציגות למשתמש את התוצאות.

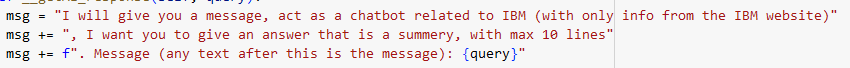
|  |  |  |
| --- | --- | --- |
| **SearchEngineUI** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | queryService: QueryService, history\_service: SearchHistoryUI | אתחול עם שירות שאילתות והיסטוריה. |
| display | אין | מציג את ממשק המשתמש. |
| perform\_search | b | מבצע חיפוש ומציג תוצאות. |

|  |  |  |
| --- | --- | --- |
| **SearchHistoryUI** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | אין | מאתחלת את ממשק המשתמש לצפייה בהיסטוריית החיפושים, כולל מידע טקסטואלי ותיבת פלט ריקה. |
| save\_search | Query: str,  results [{}] | שומרת שאילתת חיפוש ותוצאותיה להיסטוריה, שומרת רק את 5 החיפושים האחרונים ומעדכנת את התצוגה. |
| display\_history | אין | מציגה את חמשת החיפושים האחרונים (לפי הסדר), עם מספר תוצאות לכל חיפוש, בתצוגה אינטראקטיבית. |
| get\_output\_widget | אין | מחזירה את ה-Widget של היסטוריית החיפושים שניתן לשלב בטאב או ממשק אחר. |

|  |  |  |
| --- | --- | --- |
| **ChatbotUI** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | indexService: IndexService, | מאתחלת את האובייקט, כולל קונפיגורציה של מודל ה-AI והכנה של ממשק המשתמש הגרפי |
| display | אין | מציגה את ממשק המשתמש הגרפי של הצ'אטבוט. |

הערות:

* הצ'טבוט משתמש ב-API של GenAI על מנת לבנות תשובה להודעה של המשתמש, עם prompt שמדריך את ה-AI שהוא אמור לענות בתור צ'טבוט של אתר IBM ושהוא צריך לספק תשובה קצרה במקסימום 10 שורות. ה-prompt הוא:



|  |  |  |
| --- | --- | --- |
| **EditIndexUI** | | |
| פונקציה | פרמטרים | תיאור |
| \_\_init\_\_ | indexService: IndexService, crawlerService: CrawlerService, graphService: GraphService | אתחול עם שירותי אינדקס, זחילה וגרפים. |
| display | אין | מציג את התפריט לעריכת אינדקס. |