



Phoenix

ישראל אוחיון
כפיר אמויאל
שחר ברנסון
ולדי טרינטר
עומר גולדשטיין
שלומי פרידמן

מטרת מנוע החיפוש שבנינו

צד משתמש:

1. חיפוש מידע ממוקד באתר IBM עם GUI מינימליסטי להפחתת עומס בעין.
2. עם שמירת 5 חיפושים אחרונים.
3. שיחה עם ChatBot מבוסס gemini-1.5-flash לבירור מושגים ב-IBM.
4. גרף heat map לבירור קורולציה בין זוג מילים בדפים השונים.
5. גרף bars הופעת המילים מהאינדקס בדפים השונים

צד מנהל:

1. הדפסת האינדקס
2. הוספת מילה לאינדקס
3. הסרת מילה מהאינדקס
4. הוספת URL לגירוד מידע ב-crawl
5. מחיקה של URL מה-crawl
6. שמירת כל השינויים באינדקס שבמסד נתונים
7. טעינת המידע של האינדקס מהמסד נתונים.
8. ביצוע crawling מחדש בעקבות שינויים לוקליים.

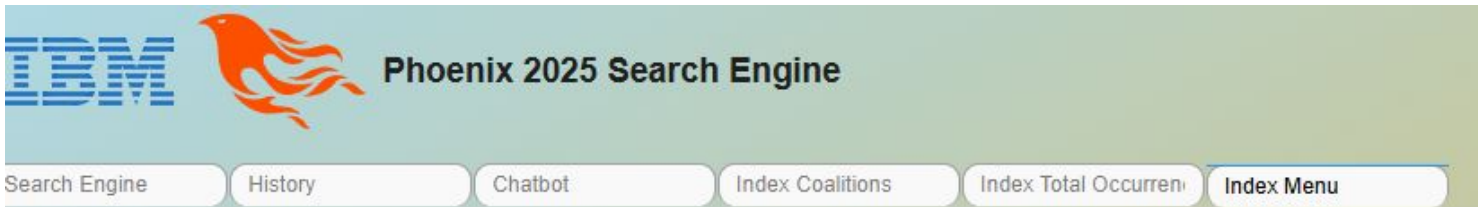
cool feature

קביעת נושא השיחה בChatBot רק לנושאים של IBM

```
74 def __getAI_response(self, query):
75     msg = "I will give you a message, act as a chatbot related to IBM (with only info from the IBM website)"
76     msg += ", I want you to give an answer that is a summery, with max 10 lines"
77     msg += f". Message (any text after this is the message): {query}"
78     try:
79         response = self.model.generate_content(msg)
80         return response.text
81     except err:
82         return "Sorry, I didn't understand (Genai is tired)"
83
```

cool feature

שינויים לוקלים באינדקס,
בלי להטריד את השרת שלא לצורך.



Menu:

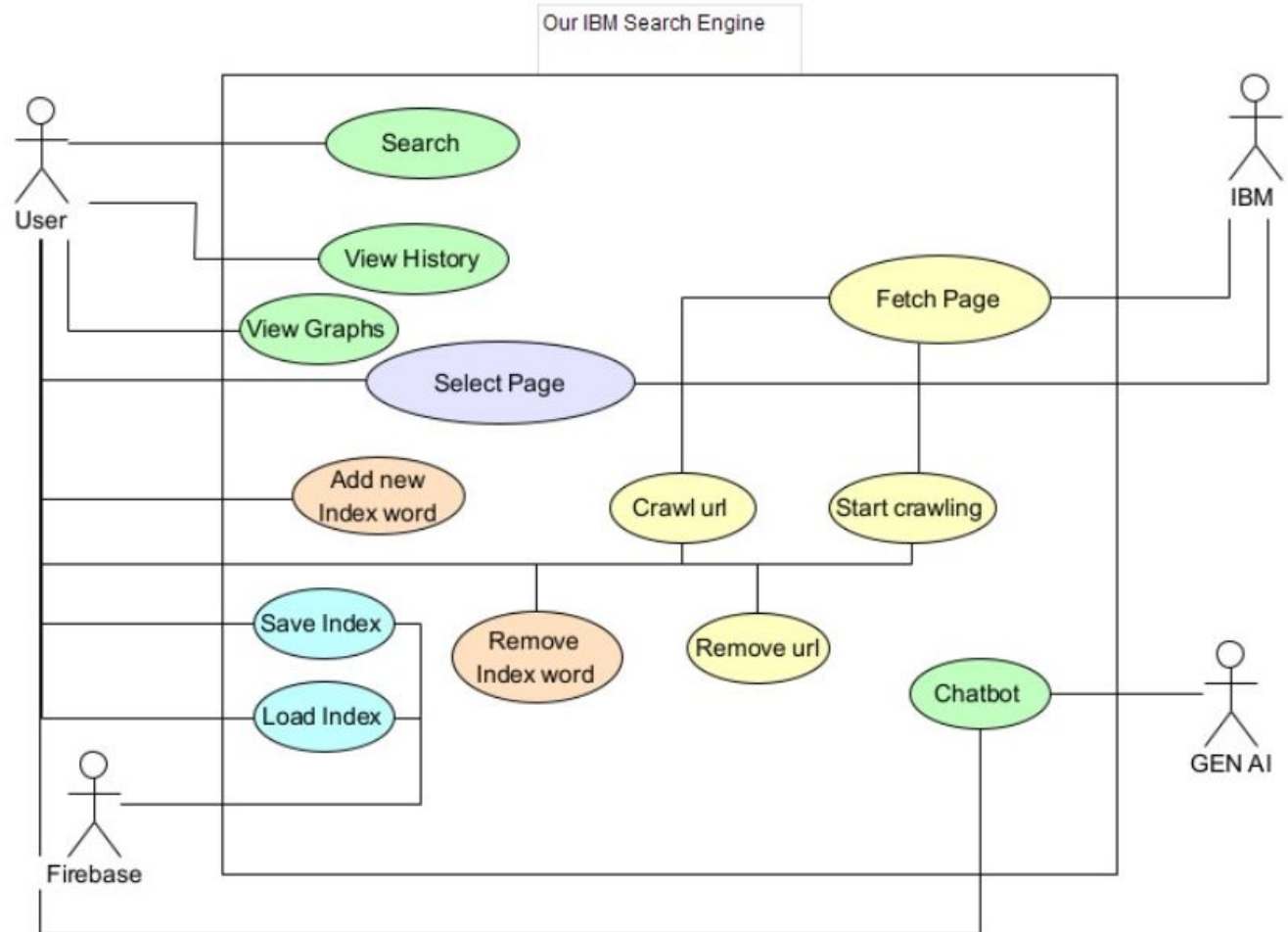
1. Print index
2. Add a new word to index
3. Remove word from index
4. Add a new url to crawl
5. Remove a url
6. Save to FireBase
7. Load from FireBase
8. Start crawling

Choose an option [1-8]:

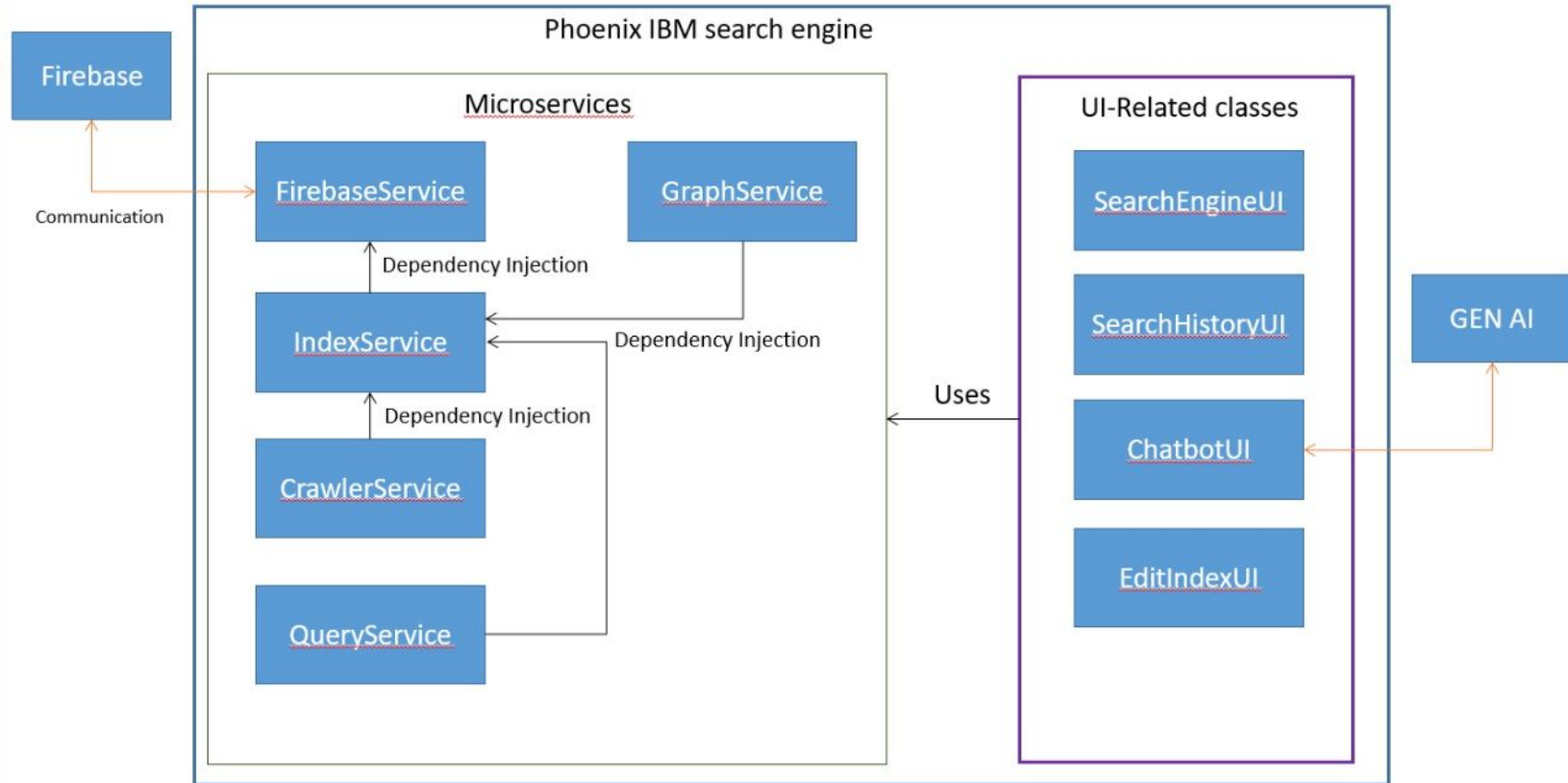
[1-8]

Send

use case



Microservices architecture



מדדי KPI לפרויקט:

שביעות רצון המשתמש:

-שהמשתמש יהיה מרוצה מתוצאת החיפוש, כלומר שהתוצאות חיפוש יכילו תוצאות נכונות.
מדידה: שתהיה לפחות תוצאה אחד נכונה \ רלוונטית ב-5 התוצאות הראשונות.

ביצועים:

-שהמשתמש לא ימתין יותר מדי זמן לתוצאת החיפוש.
מדידה: אחרי לחיצה על כפתור החיפוש לא יעברו יותר מ-3 שניות עד שתוצג תוצאה כלשהי.

מדרגיות וגמישות:

שהמערכת תהיה מודולרית, וקלה להרחבה.
מדידה: פיתוח מודולרי, שכל פונקציה ומחלקה תהיה עצמאית, כלומר בלי תלות במשתנים גלובלים.

דרישות פונקציונליות:

דרישה	סוג (FR)
המערכת מאפשרת לבצע חיפוש מידע	FR
המערכת מאפשרת זיהוי שגיאות טקסט	FR
המערכת מאפשרת השלמת טקסט	FR
המערכת מאפשרת שמירת היסטוריית חיפוש	FR
המערכת מאפשרת סיווג תכנים	FR

דרישות לא פונקציונליות:

דרישה	סוג (NFR)
Accessibility - מנוע חיפוש יוצג כשורה טקסט חופשי למשתמש.	NFR
Usability - החיפוש נעשה לפי טקסט עבור המידע באתר של IBM	NFR
Interoperability - מנוע החיפוש יהיה מתואם לאתר של IBM	NFR
Usability - היסטוריית החיפוש תהיה 5 החיפושים האחרונים של המשתמש.	NFR
Usability - הסיווג יהיה לכל תוכן הקשור לענן והמושגים באתר של IBM לפי מילות מפתח. לצורך חיפוש מהיר.	NFR
Compliance - העיצוב של המנוע חיפוש יעמוד ב-8 העקרונות של שניידרמן	NFR

ביקורת עמיתים מהסדנא:

הערת משוב	האם התבצע שינוי באפליקציה בעקבות ההערה?	נימוק
לשפר את העיצוב	כן	למרות שבדרישות שלנו יש דרישה לעיצוב מינימליסטי נשפר את העיצוב (כגון צבעים, כפתורים וכו') עבור המוצר הסופי.
זמן תפעול מהיר יותר \ יותר פירוט	כן	נוסיף טקסט שלמשתמש יהיה יותר ברור איך המערכת עובדת.
הכנסת כמות דפים ל-crawl כפרמטר	לא	כבר יש פיצ'ר להוספת דף חדש במסך עריכת האינדקס.
הצגת האינדקסים במערכת בפלט החיפוש	לא	אנו לא רוצים לחשוף את האינדקס שלנו במסך החיפוש, על מנת שלא יהיה עומס קוגניטיבי על המשתמש, המשתמש יוכל לראות את האינדקס במסך עריכת האינדקס, ולקבל מידע עליו שם ובמסכי סטטיסטיקה.
זמני ריצת החיפוש, אולי כדאי לייעל את החיפוש מראש ולשמור DB	לא	כבר מוצג למשתמש זמן החיפוש במסך התוצאות, ואנו לא רוצים לשמור את התוצאות בבסיס נתונים כדי למנוע עומס (לאחר בדיקה ה-Firebase לא אוהב את זה) בנוסף האינדקס יוכל להתעדכן בהמשך (בין במילים או בדפים חדשים \ נוספים) ולכן אנו מעדיפים שעבור כל שאלתה יהיה חיפוש חדש ועדכני
יש לשפר את האינדקס חיפוש	כן	נוסיף מילים חדשות לאינדקס
הגדל הכתב	לא	בדפדפן יש אפשרות מובנת להגדלת הכתב, לכן אין צורך לעשות את זה במערכת עצמה לאחר בדיקה עם חברי הצוות וסטודנטים אחרים, נמצא שגודל הכתב הנוכחי נוח וקל לעין
הגרף עם המספרים היה קצת מסורבל ללא הסבר מפורט מהאחראי של האתר	כן	נוסיף הסבר לגרפים

קבלנו ציון SUS: 77.64 מתוך מאה.

שקיפות אלגוריתמית

יש הסבר למשתמש לפי מה מתבצע דירוג תוצאות הדפים.
וגם הבהרה שהמנוע חיפוש שלנו תומך באופרטורים של OR ו- AND לדיוק
התוצאה.



Phoenix 2025 Search Engine

Search Engine

History

Chatbot

Index Coalitions

Index Total Occurrence

Index Menu

To query please enter text into the field and press the 'Search' button
Each page in the result is ranked by the amount the query's keywords appear in the page's text
Search for terms learned in class (IBM, AI, SAAS, etc...)
There is support for AND and OR operators

Query:

Search

שקיפות אלגוריתמית

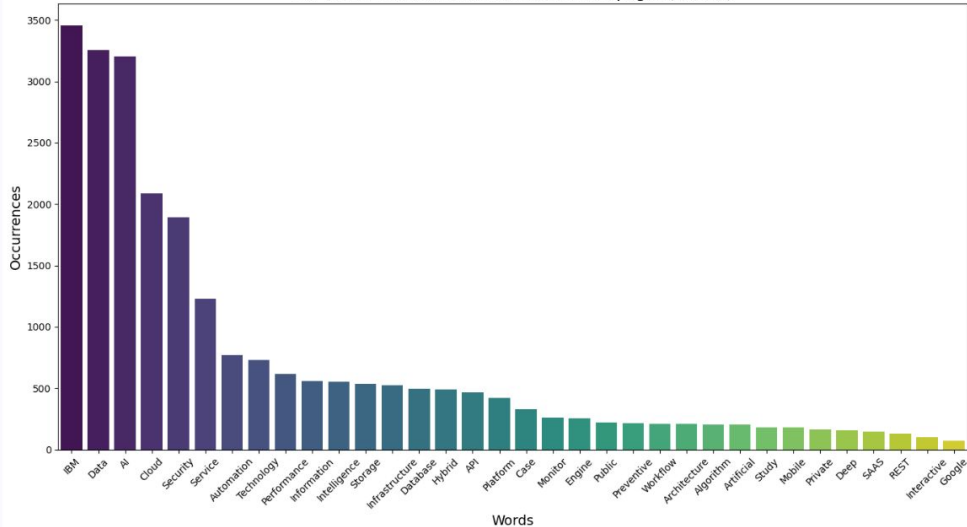
גרף ה-BAR עוזר למשתמש להבין איך מופעי המילים יכולים לבוא לידי ביטוי בדירוג התוצאות.
גרף ה-Heat Map מרמז למשתמש על היתכנות לקרולציה לינארית בין זוג מילים (כמות הדפים

ראיונות

Bar Chart: Word Occurrences Across Webpages

The value of index i is the total amount of occurrences that index showed up in the pages

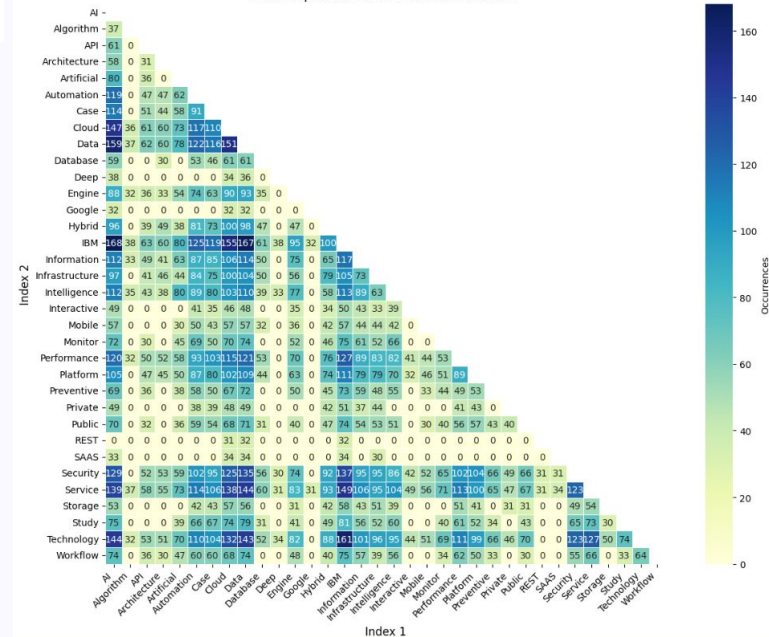
Bar Chart: Word Occurrences Across Webpages (Sorted)



Heatmap: Shared URLs between Indexes

The value at each cell $[i,j]$ is the amount of shared urls between index i and index j

Heatmap: Shared URLs between Indexes



שקיפות אלגוריתמית

1. כמות תוצאות הדפים שנמצאו.

2. סך כל הזמן שלקח להחזיר תשובה עבור השאילתא.

3. אחוז דיוק ההתאמה עבור כל דף תוצאה

(במתודה rank_url מחזירה ערך בין 0 ל-1 את התוצאה מחסרים מ-1 ומכפילים ב-100 על מנת לקבל אחוז בין 0

Query: SAAS

Search

Number of results: 34 (took 0.033 sec)

1. <https://www.ibm.com/topics/saas> (match: 98.18%)

What Is Software as a Service (SaaS)? | IBM What is software as a service (SaaS)? Cloud Link copied What is SaaS? Software...

2. <https://www.ibm.com/topics/cloud-computing> (match: 94.12%)

What Is Cloud Computing? | IBM What is cloud computing? Cloud 14 February 2024 Link copied Authors Stephanie Susnjara IBM Think Content Contributor...

3. <https://www.ibm.com/topics/paas> (match: 90.91%)

What Is Platform as a Service (PaaS)? | IBM What is platform as a service (PaaS)? Cloud 10 September 2024 Link copied Authors...

4. <https://www.ibm.com/topics/what-is-a-cmms> (match: 87.50%)

What Is a CMMS? | IBM What is a CMMS? Business operations 05 July 2021 Link copied What is a CMMS? A computerized...

5. <https://www.ibm.com/case-studies/compuware-systems-hardware-devops-z-mainframe> (match: 80.00%)

Compuware | IBM Home Case Studies Compuware Compuware Putting IBM enterprise servers at the heart of a DevOps culture Learn more As it...

6. <https://www.ibm.com/database?lnk=ProdC> (match: 80.00%)

Database Solutions | IBM Home Database Database Solutions Run your applications, analytics, and generative AI with databases on any cloud. See solutions Why...

7. <https://www.ibm.com/topics/vpc> (match: 66.67%)

What Is a Virtual Private Cloud (VPC)? | IBM What is a virtual private cloud? Compute and servers 17 June 2024 Link copied...

8. <https://www.ibm.com/topics/virtualization> (match: 66.67%)

What Is Virtualization? | IBM What is virtualization? IT infrastructure 21 March 2023 Link copied What is virtualization? Virtualization is a process that...

9. <https://www.ibm.com/topics/hybrid-cloud> (match: 66.67%)

What is Hybrid Cloud? | IBM What is hybrid cloud? Cloud 8 October 2024 Link copied Authors Stephanie Susnjara IBM Think Content Contributor...

10. <https://www.ibm.com/topics/ddos> (match: 50.00%)

What Is a DDoS Attack? | IBM What is a distributed denial of service (DDoS) attack? Security Link copied What is a DDoS...

