

קובץ דוקומנטציה:

אופן הפתרון:

- התמקדתי בכל מערכת בנפרד,
(1) תחילה בניתי מערכת המטפלת בהורדת חומרי גלם מאתר החדשות bbc ובחיפוש בתוכן הכתבות שהורדו.
(2) לאחר מכן התמקדתי במערכת המורידה נתונים על הטיסות העדכניות ביותר לפי האתר של נתב"ג, המראות ונחיתות. ובנוסף אפשרות לחיפוש לפי טקסט בפרטי הטיסות השונות.

***את שני המערכות התחלתי לבנות בגישת bottom up design, התמקדתי תחילה במימוש הפעולות הפשוטות ולאחר מכן ליכדתי אותן יחדיו ליצור את המודל המבוקש.**

Web scrape bot:

מחלקה אשר משתמשת ב webdriver של selenium ומאפשרת גישה לנתונים באתרי אינטרנט, משומשת על ידי BBC_web_scrape_bot, Flights_web_scrape_bot

BBC:

class המרכזית במערכת הינה BBC_web_scrape_bot אשר מממשת את הממשק BBC_BOT.
הפונקציות המרכזיות בממשק זה הינן :
: update_articles_storage
פונקציה אשר מורידה כתבות מהאתר ומעדכנת את מלאי הכתבות
:search_in_articles(search_text)
פונקציה אשר מחזירה את הכתבות אשר מכילות את המילים מטקסט החיפוש

BBC web scrape bot

מכילה פונקציות אלו וממשת אותן בדרך הבאה:
:update articles storage
הפונקציה משתמשת ב Web_scrape_bot ומושכת מידע על כל הכתבות בדף הראשי של bbc.
מושכת את הלינקים של כל כתבה, לאחר מכן עוברת על כל לינק, פותחת לינק זה ומושכת את תוכן הכתבה, לבסוף שומרת את פרטי הכתבה ב json בתיקייה המתאימה.
:search in articles(search text)
הפונקציה מקבלת קלט של טקסט מפרידה אותו למילים, ולאחר מכן עוברת על כל קבצי ה Json בתיקייה המתאימה ובודקת בתוכן של הכתבות, האם לפחות חצי מהמילים מופיעות בתוכן. כל הכתבות שעונות על התנאים מודפסות למסך בטבלה.

מבנה השימוש במחלקה:

```
BBC_web_scrape_bot(BBC_BOT)
Web_scrape_bot
Links_extract
Article_content_extract
Article_information
Article_search
```

:Links_extract

מחלקה העוברת של שני סוגי אלמנטים מן הדף הראשי של bbc ומוציאה את הלינק של כל אחד מהכתבות לפי האלמנט שבו נמצאת, לאחר מכן מחזירה את הרשימה עם הלינקים.

:Article_content_extract

תחילה מורידה משם הכתבה אותיות העלולות לגרום לבעיה ביצירת הקובץ (כגון: *, ?, ...) לאחר מכן בודקת האם קיים קובץ json לכתבה זו, אם לא בודקת את סוג הכתבה ומבצעת משיכת נתונים מהאלמנט בהתאם לסוג. לבסוף יוצרת קובץ Json בתיקייה המתאימה. את פרטי הכתבה שומרת בArticle_information

:Article_information

מחלקה אשר מחזיקה את הנתונים על כתבה : שם, לינק, תוכן

: Article_search:

מקבלת כקלט מחרוזת, מפרידה אותה למילים. עוברת על כל קבצי json של הכתבות, ואם לפחות חצי מהמילים מוכלות בתוכן כתבה, אזי הכתבה נשמרת ברשימה. לבסוף מודפסת הרשימה של כל הכתבות שעונות על התנאים.

דוגמא:

דוגמא להדפסת טבלה של כתבות אשר מכילות את המילה "Israel":

Article Name	Article Link
Bsissa North Africa's ancient convenience food - BBC Travel	https://www.bbc.com/travel/article/20220130-bsissa-north-africas-newest-brea
Israeli policies against Palestinians amount to apartheid - Amnesty - BBC News	https://www.bbc.com/news/world-middle-east-60197918
Survivors and leaders mark Holocaust Memorial Day - BBC News	https://www.bbc.com/news/in-pictures-60154892
True story Lie detection systems go high-tech - BBC News	https://www.bbc.com/news/business-60153129
Why modern food lost its nutrients	https://www.bbc.com/future/bspoke/follow-the-food/why-modern-food-lost-its

Flights:

הclass המרכזית במערכת הינה **Flights_web_scrape_bot** אשר מממשת את הממשק **FLIGHTS_BOT**. הפונקציות המרכזיות בממשק זה הינן:

:update_flight_storage

פונקציה אשר מורידה את פרטי הטיסות העדכניות מהאתר

:search_in_flights(search_text)

פונקציה אשר מחזירה את הטיסות אשר מכילות את המילים מטקסט החיפוש

: Flights web scrape bot

מכילה פונקציות אלו ומממשת אותן בדרך הבאה:

:update flight storage

הפונקציה משתמשת ב**Web_scrape_bot**, הוצאת המידע מהאתר תתבצע רק אם האובייקט מכיל את הגרסה החדשה ביותר של האתר.

הפונקציה מושכת מידע על כל הטיסות העדכניות בדף של נתב"ג, תחילה יורדת עד סוף הדף, ומבצעת לחיצה על ביטול עדכון אוטומטי, זאת כדי שכאשר מתבצע הורדה של מידע מאלמנטים בדף, לא נרצה שאלמנטים אלו ישתנו בזמן הוצאת המידע.

מושכת את פרטי כל הטיסות הנכנסות ומאכסנת מידע זה כ**json** בתיקייה המתאימה. לסיום עוברת לטבלה של טיסות יוצאות, יורדת עד סוף העמוד ומורידה את המידע על טיסות אלו, את המידע מכניסה ל **json** בתיקייה המתאימה.

:search in articles(search text)

הפונקציה מקבלת קלט של טקסט מפרידה אותו למילים, מבצעת עדכון פרטי טיסות, ולאחר מכן עוברת על כל קבצי ה **Json** בתיקיות המתאימות ובודקת בתוכן של פרטי הטיסה, האם מילה מן הטקסט מופיעה באחד מפרטי הטיסה. כל הטיסות שעונות על התנאים מודפסות למסך בטבלאות.

מבנה השימוש במחלקה:

Flights_web_scrape_bot(FLIGHTS_BOT)

Web_scrape_bot

Extract_flights_table

Extract_and_update_flights

Flight_arriv_information

Flight_depart_information

Update_flights_storage

Flights_search

:Extract_flights_table

המחלקה עוברת על הדף של נתב"ג יורדת עד סופו , מבטלת את העדכון האוטומטי ומושכת את הטבלה של הנחיתות, לאחר מכן עוברת למיקום של הטבלה של המראות ומושכת את הנתונים ממנה

:Extract_and_update_flights

תחילה בודקת איזה סוג מידע של טיסות קיבלה לפונקציה, מוחקת את הקבצי טיסה הישנים מהתיקיה לפי סוג טיסה. לאחר מכן מושכת את המידע מהאלמנטים ומכניסה למבנה נתונים המתאים . אם הטיסות בנחיתות אז שומרת את פרטי הטיסה במחלקה מסוג Flight_arriv_information , אם הטיסות בהמראות שומרת את פרטי הטיסה ב Flight_depart_information . שומרת את כל הטיסות ברשימה ולאחר מכן לפי הרשימה יוצרת קבצי json של הטיסות לפי סוג הטיסה.

:Flight_arriv_information

מחלקה אשר מחזיקה את הנתונים על טיסה בנחיתות : חברת תעופה, מספר טיסה, נוחת מ , טרמינל, זמן מתוכנן, זמן עדכני, סטאטוס

:Flight_depart_information

מחלקה אשר מחזיקה את הנתונים על טיסה בהמראות : חברת תעופה, מספר טיסה, נוחת מ , טרמינל, זמן מתוכנן, זמן עדכני, סטאטוס, דלפק

:Update_flights_storage

יוצרת json בהתאם לסוג הטיסה ושומרת אותו בתיקיה המתאימה

:Flights_search:

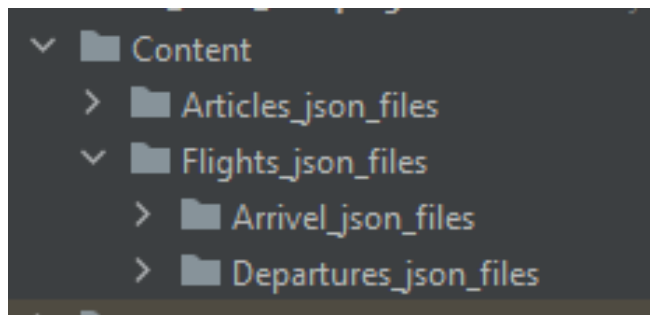
מקבלת כקלט מחרוזת, מפרידה אותה למילים . עוברת על כל קבצי הjson של הטיסות, ואם לפחות אחת מן המילים מוכלות בתוכן אחד השדות של הטיסה, אזי ההטיסה נשמרת ברשימה. לבסוף מודפסת הרשימה של כל הטיסות שעונות על התנאים בשני טבלאות לפי סוג טיסה,

דוגמא:

דוגמא לחיפוש של המחרוזת "America"

ARRIVAL FLIGHTS							
airline name	land_from	status	flight_num	terminal	schedule_time	updated_time	
american airlines	ניו יורק	לא סופי	AA 146	3	17:15 02/02	17:15	
american airlines	לונדון	בנחיתה	AA 8372	3	21:05 01/02	20:39	
american airlines	פריס	לא סופי	AA 8376	3	20:00 02/02	20:00	
american airlines	מילאנו	לא סופי	AA 8386	3	16:00 02/02	16:00	
american airlines	ברצלונה	סופי	AA 8390	3	19:40 01/02	22:12	
DEPARTURE FLIGHTS							
airline name	land_from	status	flight_num	terminal	schedule_time	updated_time	counter
american airlines	ניו יורק	בזמן	AA 145	3	00:45 02/02	00:45	56-62
american airlines	לונדון	בזמן	AA 8371	3	09:05 02/02	09:05	78-99
american airlines	מילאנו	בזמן	AA 8385	3	06:50 02/02	06:50	78-99

דוגמא למבנה התיקיות המכילות את קבצי הJson השונים בפרויקט:



* הסברים ארוכים ומפורטים יותר על כל המחלקות פונקציות ומשתנים ניתן למצוא בהערות בקוד