אוטומציה - WEB

**מטרת הפרויקט:**

לבנות מערכת המקבלת כפרמטר ממשתמש לינק לאתר

(אפשר להעביר את הלינק גם כפרמטר בהפעלת התוכנית , command line param)

ובודקת כמה אלמנטים של HTML יש לעומת כמה צריך להיות

הדרישה של כמה צריך להיות מוגדרת כטבלה בבסיס נתונים

בבסיס הנתונים שלנו יש טבלה המכילה את שם האתר ואת האלמנטים השלו

כאשר סורקים אתר מסויים , מתייחסים רק לאלמנטים של אותו אתר

דוגמא :

אתר : www.mySite.com

עבור אתר שלי הHTML ניראה ככה:

<div class="wmx8 mx-auto grid jc-center" role="alertdialog" aria-describedby="notice-message">

<div class="grid--cell mr12" aria-label="notice-message">

<p class="mb0 lh-lg">

By using our site, you acknowledge that you have read and understand our <a class="s-link s-link\_\_inherit td-underline fc-white" target="\_blank" href="https://stackoverflow.com/legal/cookie-policy">Cookie Policy</a>, <a class="s-link s-link\_\_inherit td-underline fc-white" target="\_blank" href="https://stackoverflow.com/legal/privacy-policy">Privacy Policy</a>, and our <a class="s-link s-link\_\_inherit td-underline fc-white" target="\_blank" href="https://stackoverflow.com/legal/terms-of-service/public">Terms of Service</a>.

</p>

</div>

<a class="grid--cell fc-white js-notice-close" aria-label="notice-dismiss">

<svg aria-hidden="true" class="svg-icon m0 iconClear" width="18" height="18" viewBox="0 0 18 18"><path d="M15 4"></path></svg>

</a>

</div>

הטבלה נראית ככה : (אחרי פילטור לפי שם האתר)

|  |  |  |
| --- | --- | --- |
| **Valid number** | **element** | **site\_name** |
| 10 | div | www.mySite.com |
| 3 | span | www.mySite.com |
| 3 | a | www.mySite.com |
| 2 | p | www.mySite.com |

הפלט יהיה :

P expected 2 found 2

a expected 3 found 3

Span - expected 3 found 0

Div - expected 10 found 2

**הכנה לפרויקט והערות :**

* הפרויקט הוא אמיתי גדול מסובך ומכיל הרבה מרכיבים
* אין ציפייה שתסיימי תוך שבוע- אבל אני כן ארצה לראות התקדמות, ככה שאת הגיט יש לשלוח לי על ההתחלה ברגע שיש לפחות משהו רשום. כן חשוב להכניס קוד לגיט כול כמה זמן ולא לשפוך הכול בסוף
* לא לפחד להשתמש בקוד מהאינטרנט או מספריות מוכנות **אבל** חשוב מאוד להבין מה העתקת גם אם זה עובד טוב!

**התקנות :**

צריך להתקין איזה שהוא בסיס נותנים שתוכלי לעבוד מולו. אם כבר עבדת אם משהו בעבר מה טוב אם לא תנסי להתקין mySQL

**המשימה:**

* המישמה הפעם לא שבורה למשימות קטנות , אלא יותר לשלבים מנחים של דרך הפעולה.
* כמובן שאת מוזמנת לעשות את זה בדרך שלך העיקר שבסוף זה יעבוד!
* תנסי לשבור את הפרויקט לקבצים לפי נושאים . עבודה מול בסיס נותנים , עבודה מול האתר, חשובים ובדיקות

**שלב 1**  - התחברות לביסיס נתונים:

* תיצרי טבלה אחת בודדה בבסיס נתונים שלך
* תמלאי אותה באיזה שהם ערכים
* תנסי להתחבר לבסיס נתונים שלך מהקוד ולשלוף מתוך הטבלה את הנתונים שהרגע הכנסת
* טיפ:

Each DB usually has : url , port, user , psw and database name

In order to connect to the DB you need them all .

Here is an example how to connect to mySQL from Python and run sql query there :

import MySQLdb   
# Connect  
 db = MySQLdb.connect(host="localhost", user="appuser", passwd="", db="onco")   
cursor = db.cursor()   
# Execute SQL select statement   
cursor.execute("SELECT \* FROM location")

* ברגע ההתחברות הצליחה. ניתן לייצר טבלה אמיתית עם נתונים שאנחנו מצפים לקבל מהאתר ולנסות לכתוב שאילתא שמביאה את הנתונים האלה

**שלב 2**  -קבלת דף ה HTML של האתר :

* סריקת אתר אינטרנט נקראת : scraping באנגלית , וזה מה שאתם מחפשים באינטרנט
* כול קוד שתימצאו ולא יהיה מסובך מדי , הוא לגיטימי לצורך הפעולה הזאת
* את הקוד יש לשים בקובץ פייטון נפרד ולהפעיל רק פונקציות שלו
* שימו לב: יש אתרים שחוסמים scraping , לא להילחץ מזה , פשוט לנסות לשים אתר אחר (URL אחר)
* התוצאה של הפונקציה של הסיקה צריכה להחזיר טקסט שהוא בעצם HTML

**שלב 3**  מעבר על הדף וביצוע החישוב :

* ברגע שיש לנו את הנתונים שמפים לקבל משלב 1 ואת האתר עצמו משלב 2
* אנחנו יכולים לכתוב קוד שמחפש אלמנטים בתוך הטקסט ומדפיס כמה נמצאו לעומת כמה נדרש

**שלב 4**  קבלת שם אתר מהמשתמש

* לרשום פונקציה שמקבלת את שם האתר מהמשתמש מריצה עליו את שלב 3 ומדפיסה פלט
* את הפעולות בראשיות יש להכניס לפונקציית \_\_main\_\_

רק היא מופעלת בעת הפעלת התוכנית. היא מפעילה את כול הפעולות האחרות

לא לשכוח לעשות ולידציות

לבדוק מקרי קצה (אתר לא זמין, חזר ריק, אין נתונים בבסיס נתונים וכו )

בהצלחה!

דינה