

המחלקה להנדסת תוכנה

פרויקט גמר – תשפ"א

אוטומציה באתרים

Website automation

מאת

אלישבע דיין תז: 208027102

תאריך: 02/05/2021



מנחה אקדמי: מר אלי איזק אישור:

תאריך:

רכז הפרויקטים דר' אסף שפנייר אישור:

מערכות ניהול הפרויקט:

#	מער כת	מיקום
1	מאג ר קוד	https://github.com/elishevada/final_project.git
2	יומן	https://trello.com/invite/b/xrdCfugs/87361ba5a5ca4f99d954f8e0ba05ddff/wa
5	סרט ון גיסר ת אלפ א	https://drive.google.com/file/d/11tysiYmB_3HWhUe8TZjCnB9VmciRsBow/view?usp=sharing

מידע נוסף:

סוג הפרויקט	2-יוזמה של מנחה
פרויקט מח"ר	כן
פרויקט ממשיך	פרויקט חדש

מילון:

תג: מאחורי כל דף אינטרנט ישנו קוד html המייצג דף זה, כל ארגומנט בדף מיוצג ע"י תג.

0. נאום המעלית

הרעיון של אוטומציה באתרים זה שפעולות יעשו בצורה אוטומטית מבלי התערבות אנושית .
הוא נותן מענה על אורך הזמן שלוקח לבצע פעולות על אלמנטים שונים.
הוא מיועד לכלל האוכלוסיה ביניהם מרצים סטודנטים אנשי המדיה החברתית וכו.

1. מבוא

במסגרת הפרויקט נפתח אלגוריתם שמבצע אוטומציה באתרי אינטרנט שונים.
מטרת הפיתוח היא להקל על משתמשי אתרים שמבצעים אותם פעולות כמה פעמים ביום או שצריכים להזין נתונים וכל פעם לבצע את אותם פעולות על כל נתון .
פיתוח האלגוריתם יאפשר למשתמשים בו להזין שם אתר כלשהוא שהם רוצים שיתבצע האוטומציה ושירות מבוקש וע"י זה נציג לו את מה שביקש.

2. תיאור הבעיה

אפיון הבעיה :
על מנת לבצע פעולות לקבל מידע או קובץ באתרי אינטרנט נדרשים לעשות הרבה פעולות כדי להגיע למוצר אותו רוצים לקבל.
מטרת הפרויקט היא לאפשר למשתמש לקבל את מה שהוא רוצה תוך מספר פעולות מינימליות הדרושות לצורך חיפוש ואימות.

דרישות :נספחים

אתגרים טכנולוגיים:

- מציאת אלמנטים רבים בדף ה html של האתר ולפעמים הם אותו סוג תג מה שמקשה על חיפושם.
- ישנם אתרים שבהם האוטומציה לא תתבצע בעיקבות הגנה מפני תוכנות אוטומטיות ורובוטים .
- רוצים שהאלגוריתם יעבוד על כמות גדולה של אתרים והכל בצורה כללית כך שלא נצטרך לשנות לכל אתר את האלגוריתם בהתאם לכמות הקלט אותו צריך להזין.
- ביצועים- האלגוריתם צריך לספק את השרות המתבקש במהירות סבירה .
- אינטגרציה-הטמעת האלגוריתם בסביבת הלקוח.

3. תיאור הפתרון

סקירה :

בשלב ראשון נעסוק בעיקר בניסוי על כמה אתרים בודדים, בבדיקת הנתונים המוזנים ע"י המשתמש והזנתם באתר, ונחפש את השירות המבוקש באתר ע"י שימוש בתגי html ונוציא אותו למשתמש . בשלב השני נלמד את האפשרויות השונות של הכנסת קלט לאתר ונרצה למקסם את היכולות של האלגוריתם כך שיהיה עבור כמות גדולה יותר של אתרים ובאופן כללי יותר.

תיאור הפתרון:

שלב א :

בשלב זה נגדיר את הצעדים הבאים:

א. בדיקה

- נבקש קלט מהמשתמש של שם אתר ממנו הוא רוצה את השירות.
- בדיקת נכונות של הכתובת שהתקבלה.

ב. ביצוע

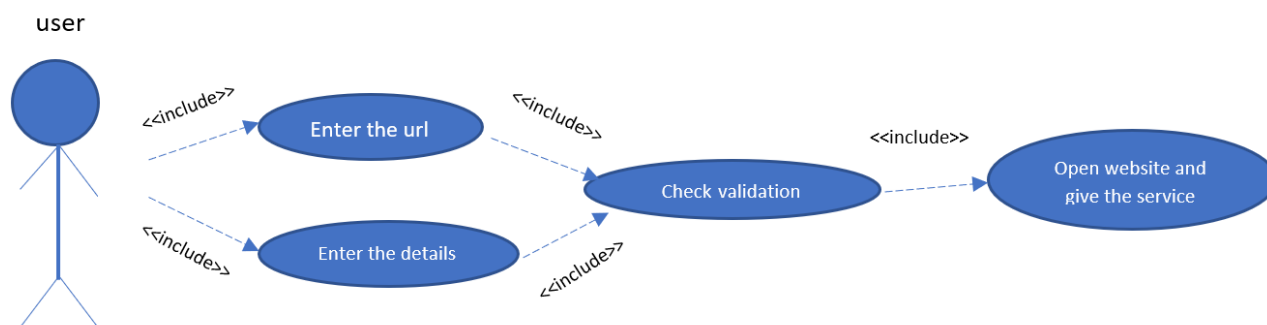
- נחפש את השירות המבוקש ע"י חיפוש אלמנטים בדף ה-html ע"י פונקציות `find_element` של ספריית פייטון .
- נגיע לשירות המבוקש ונוציא אותו למשתמש.
- ג. הרחבת האלגוריתם
- נבין את סוג האתר אם זה אתר שצריך לחפש בו מידע כמו פייסבוק או אתר שצריך להוריד ממנו קובץ כלשהוא כמו ביטוח לאומי , או אתרים שצריך בהם סיסמא .
- נרצה להגדיל את השרות עבור מספר גדול יותר של אתרים ובצורה כללית יותר.
- נרצה גם אפשרות של קריאה מקובץ נתונים כך שהמשתמש יכול להכניס קובץ של נתונים ונקרא את הנתונים ונכניסם לאתר, זה במקרה שצריך לבצע את אותם פעולות על הרבה נתונים .
- כמו למשל עדכון באתר המכללה של שם הסטודנט ותז , נרצה לעשות זאת עבור הרבה תלמידים.

כלים ועזרים נוספים:

- כלי אוטומציה : selenium
- שימוש בספריות פייטון על אתרים בchrome

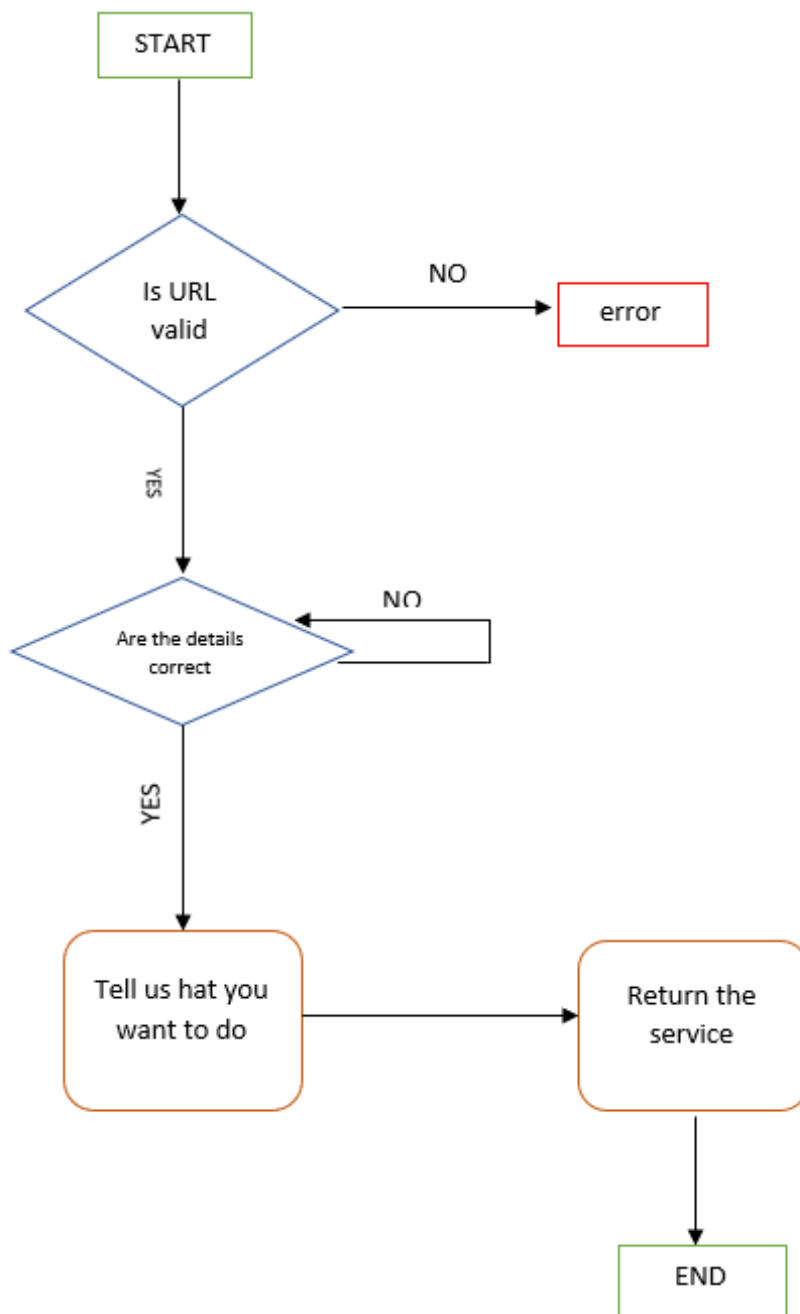
Use Case Diagram:

התרשים הבא מתאר את ארכיטקטורת המערכת.
יש לנו משתמש שצריך לשים כקלט קישור חוקי לאתר כלשהו שבו הוא רוצה שיתבצע האוטומציה.
לאחר מכן יש בדיקה של הקלט באם הוא נכון אפשר להמשיך לשלב שבו מבקשים את פרטי המשתמש, בודקים את נכונותם ואחר כך עוברים לשלב שמחזיר את התשובה או קובץ כלשהוא או מידע וכו'.





תרשים זרימה:



4. תוכנית בדיקות

בדיקות קלט:

- בדיקה האם url תקין ופרטים תקינים.

בדיקות יחידה :

- בדיקה לאחר כל הוספת פיצרים האם התכנית עובדת כמצופה .

בדיקות מערכת:

- בדיקת Flow מקצה לקצה ע"י בדיקת פעולה על אתר שתוצאתה ידועה מראש.
- בדיקת תפקוד האלגוריתם בזמן אמת (מבחינת ביצועים וכיוצ"ב).

בדיקות אינטגרציה:

- בדיקת פעולת המערכת לאחר הטמעת האלגוריתם.
- בדיקת ביצועים של המערכת השלמה.

5. סיכום \ מסקנות

סיכום:

עד כה עסקנו בקבלת קישור לאתר כלשהוא ובדיקה איזה פרמטרים הוא צריך, הזנו את הנתונים הנדרשים וקיבלנו שירות כלשהוא שהגדרנו במערכת .

כעת נרצה להוסיף יכולות למערכת שתעבוד ליותר אתרים ולהרבה בקשות שירות ונרצה להוסיף בדיקות עבור כל הכנסה של פרט כלשהו , בנוסף נרצה לשמור את פרטי המשתמשים כך שלא יצטרכו להכנס כל פעם ולהזין מחדש את פרטיהם.

6. נספחים

טבלת סיכונים

Risk description	prob	impact	Plan to reduce impact
קריסת המשאבים הדרושים להפעלת האתר.	50%	H	ייעול הקוד.
עבודה עם כלי חדש.	100%	H	לימוד עצמי.
אלמנט ללא זיהוי הגיוני או אמצעי זיהוי ארוך סירבול הקוד.	70%	H	מציאת האלמנט ע"י תג מסוים או נתיב.
קיום מחויבויות נוספות כמו עבודה, לימודים וכו'	70%	H	ניהול זמנים בצורה נכונה ויעילה וקביעת עוגנים.

נרצה ללמוד על אתרים החסומים.	H	30%	חסימה באתר מפני רובוטים
------------------------------	---	-----	-------------------------

טבלת דרישות

❖ דרישות פונקציונליות

מס' דרישה	תיאור
1	האלגוריתם יעבוד עם מספר גדול של אתרים.
2	אם משתמש נכנס כבר למערכת נעלה את פרטיו שוב.

❖ דרישות נוספות ולא פונקציונליות

- מהירות (זמן): מהירות האלגוריתם ומתן מענה מהיר.
- עומס: במצבי עומס נרצה להוציא הודעה למשתמש על המתנה.
- זמינות: הלקוח יקבל את מה שהוא רוצה.

