

פייתון

תרגיל מסכם - logpuzzle

תרגיל logpuzzle

- ▶ תרגיל מסכם בפיתון.
- ▶ מכיל דברים שלמדנו ועוד כמה חדשים.
- ▶ המטרה: להרכיב תמונה מחלקים שלה שפוזרו ברחבי האינטרנט.

URL – Universal Resource Locator ▶

לכל משאב באינטרנט – יש URL ייחודי ▶

למשל: ▶

<http://www.blabla.com/blapage.html> ◦

<http://www.blabla.com/bla1.pdf> ◦

מהו דומין

ספריית urllib

- ▶ השימוש ב `urllib.request.urlopen` עם הפונקציה `urlopen` יאפשר לנו להוריד קובץ תמונה מהאינטרנט ולשמור אותו כקובץ תמונה במחשב שלנו.
- ▶ יש להעביר לפונקציה פרמטר שכולל את ה URL של הקובץ שרוצים להוריד ולשמור.
- ▶ הדרכה:
 - גלשו לאתר כלשהוא, בחרו תמונה ולחצו לחצו ימני.
 - ואז "פתח תמונה בכרטיסיה חדשה".
 - עיברו אל הכרטיסייה החדשה והעתיקו מהדפדפן את ה URL שלה.
 - מיד נשתמש ב - URL בקוד פיתון.

קוד פיתון להורדת תמונה מהאינטרנט

```
import urllib.request
URL = 'https://shapira.ort.org.il/wp-
content/uploads/2018/10/IMG-20200609-WA0051.jpg'
FILENAME = r'C:\Users\Zvika\Desktop\logpuzzle\image.jpg'

with urllib.request.urlopen(URL) as response:
    image = response.read()
    with open(FILENAME, 'wb') as output_file:
        output_file.write(image)
```

איך מחברים כמה תמונות לתמונה אחת?

- ▶ למדנו איך להוריד תמונה אחת מהאינטרנט.
- ▶ כדי לפתור את הפאזל – נצטרך לחבר את התמונות שנוריד מהאינטרנט לתמונה אחת.
- ▶ נשתמש בקובץ `html`.
- ▶ נרשום בתוכו את שמות התמונות שנוריד – ואז הפעלת קובץ ה-`html` (הקלקה כפולה) תציג אותן בדפדפן זו לצד זו כתמונה אחת.

תבנית לקובץ html

```
1 <html>
2 <body>
3   <img src = "images/pic1.jpg">
4   <img src = "images/pic3.png">
5 </body>
6 </html>
```


התרגיל - logpuzzle

- ▶ בקישור נתונים שני קובצי לוג של גישה לאינטרנט שמכילים טקסט ובתוכם מתחבאים url-ים של תמונות jpg שאותן יש להוריד.
- ▶ התמונות נפרסו לחתיכות אנכיות וכל URL מכיל חתיכה אנכית מתוך התמונה.
- ▶ לאחר שתורידו מהאינטרנט את קבצי התמונות – יש להסיר כפילויות ולמיין את התמונות לפי השם.
- ▶ הורידו את קבצי התמונה ושימרו אותם במחשב שלכם תחת הסירו כפילויות ומיינו את הקבצים לפי השם. ניתן לתת לכל תמונה שם פשוט –img0, img1.
- ▶ יש ליצור ע"י הקוד שלכם קובץ html ובתוכו לשים את התמונות שהורדתם עפ"י הסדר.
- ▶ שם קובץ ה- html: index_animal.html, index_place.html
- ▶ הקלקה על קובץ ה- html תציג לכם את התמונות.

הנחיות נוספות

▶ את התוכנית יש להריץ עם פרמטרים מתוך
(PyCharm) edit configurations
או מה - CMD, כך:

▶ `--todir c:\logpuzzle c:\logpuzzle\place_code.google.com`

▶ הסבר:

- `C:\logpuzzle` תיקייה אליה נשמרים קבצי התמונה שיוורדו מהאינטרנט.

- `c:\logpuzzle\animal_code.google.com` שם קובץ הלוג שבו נמצאים הקישורים לתמונות.

בתיקייה תמצאו את הקבצים הבאים:

- ארבעה קבצי לוג שאפשר לפתוח עם notepad++:

- animal_code.google.com

- place_code.google.com

- logo_data.cyber.org.il

- message_data.cyber.org.il

- [logpuzzle.py](#) – קובץ שלד – לעזרה בפיתוח הקוד.

- אתם מוזמנים להוסיף פונקציות ככל שנדרש.

► שימו לב : שם הדומין נמצא בתוך שם הקובץ

ציונים יינתנו כך:

70	התוכנית עובדת נכון, מקבלת פרמטרים, מורידה את ה URLs, יוצרת קובץ html וניתן לראות את התוצאה. יש שימוש בפרמטרים (edit configuration).
20	קוד כתוב נכון ויפה, חלוקה לפונקציות, שמות משמעותיים, קבועים, try/except
10	Pep8, תיעוד
100	סה"כ

בהצלחה!