**קובץ תיעוד לפרוייקט 1 - ניהול נתונים באינטרנט**

לירון כהן 207481268, יובל מור 209011543

**תיאור הקוד**

חלק א' - ייצור קובץ האונטולוגיה (create\_graph)

* ייצור רשימת ה-urls של המדינות מהעמוד הנתון (get\_countries\_urls)
  + הורדת תוכן העמוד הנתון באמצעות requests, lxml.
  + שימוש בשאילתת xpath (COUTNRIES\_XPATH\_QUERY) על העמוד הנתון להוצאת המדינות.
  + הוספה ידנית של שלוש מדינות באמצעות שאילתות ייעודיות.
* הוספת שלישיות לגרף עבור כל מדינה (add\_triplets\_to\_graph)
  + פרסור שם המדינה והוספתו לקבוצת המדינות.
  + הורדת תוכן עמוד המדינה באמצעות requests, lxml.
  + הוספת שלישייה לגרף באמצעות הרצת שאילתת ה-xpath המתאימה עבור   
    ה-relation המתאים (add\_country\_triplet\_to\_graph).
  + אם מדובר בשאילתת president או prime minister מתבצע מעבר לעמוד האדם והוספת השלישיות המתאימות לו (add\_person\_triplets\_to\_graph), תוך התייחסות למדינת הלידה בהשוואה לרשימת המדינות שהתקבלו מהעמוד הנתון.
  + הוספה ידנית של שאילתת population מיוחדת עבור מקרי קצה.
* יצירת האונטולוגיה ושמירתה
  + מתבצע באמצעות פקודה serialize על הגרף שהתקבל.

חלק ב' - מציאת תשובה לשאלה שנשאלה (ask\_question)

* המרת השאלה לשאילתת sparql (parse\_question\_to\_query)
  + מציאת סוג השאלה לפי מילת השאלה ומילות מפתח נוספות.
  + פרסור הפרמטרים הנחוצים לשאילתא מתוך השאלה.
  + ייצור השאילתא המתאימה (generate\_<x>\_query) והחזרתה.
* ייבוא הגרף מהקובץ והרצת השאילתא
  + פרסור המשתנה המתאים שהגיע מהשאילתא.
  + סידור התשובה ברשימה ממוינת לפי הצורך.
  + התייחסות מיוחדת לשאילתת who is שנכתבה כאיחוד של שתי שאילתות (אחת   
    ל-president ואחת ל-prime minister).
  + הדפסת התשובה ויציאה.

**תיאור השאלה שהוספנו**

השאלה שהוספנו היא:

List all countries that their names and their capitals' names end with the string <str>

דוגמאות לתשובות אפשריות

* עבור המחרוזת "ia" התקבלו התוצאות Bulgaria (שעיר בירתה Sofia) ו-Liberia (שעיר בירתה Monrovia).
* עבור המחרוזת "e" התקבלה התוצאה Zimbabwe (שעיר בירתה Harare).

**תיאור מקרי קצה**

1. חילוץ מדינות מהעמוד הראשי
2. מדינות Population
3. מדינות birth place (דוגמת ישראל להרצוג ובנט)