**הערות כלליות:**

-חלק מאתרי האינטרנט שמהם תנסו לחצוב מידע משתמשים בטכנולוגיות שבהן שימוש ב-Jsoup לא יעבוד. Jsoup היא ספרייה שמאפשרת לבצע מניפולציות על מסמך HTML קיים. הספרייה לוקחת את קוד ה-HTML מכתובת מסוימת, ולאחר מכן יכולה לבצע מניפולציות על המסמך הזה.

ישנם אתרי אינטרנט שבהם קוד ה-HTML הראשוני שנוצר הוא מינימלי, ורוב המידע מתווסף אליו רק אחרי ש-Jsoup אסף את המסמך. באתרים כאלה – לא תוכלו לחצוב מידע באמצעות Jsoup, אלא באמצעות טכניקות אחרות שנכיר בהמשך הקורס. לכן, אם אתם מגלים עבור אתר מסוים שהאלמנטים שמופיעים בדפדפן אינם מוחזרים על ידי Jsoup, ותרו על האתר הזה ועברו לאתר אחר.

-כל הפרויקטים אמורים לעבוד עם ממשק משתמש (GUI) הנכתב באמצעות הספרייה swing שאותה הכרנו בתחילת הקורס. אין להגיש פרויקטים שרצים בתוך ה-console של סביבת העבודה.

**פרויקט 1: מערכת לניטור מילות חיפוש באתרי חדשות**

תיאור כללי: המערכת תסרוק אתרי חדשות ותחפש בכתבות השונות שלהם ערך שהוזן על ידי המשתמש. כל הכתבות המתאימות יוצגו עבור המשתמש.

פירוט: עם הרצת התוכנית, ייפתח חלון (GUI) שבו יופיעו תיבת טקסט וכפתור. המשתמש יוכל להכניס מילה/ביטוי לתיבת הטקסט, ולאחר שילחץ על הכפתור, יופעלו Web Scrapers על לפחות חמישה אתרי חדשות. ה-scraperים יקראו את כל הכתבות אליהן מפנה העמוד הראשי של אתר החדשות, ועבור כל כתבה בה נמצאת התאמה לערך החיפוש, יופיע בחלון ה-GUI המידע בפורמט הבא לדוגמה:

אתר: one.co.il

כותרת הכתבה: ניצחון למכבי תל אביב במשחק העונה

קישור: <https://one.co.il/my_article>

החיפוש של הכתבות יתבצע בלולאה אינסופית. כלומר, מיד לאחר שה-scraper סיים לסרוק את האתר שעליו הוא עובד, הוא יבצע סריקה נוספת כדי לבדוק האם כתבות נוספות עם ערך החיפוש נוצרו בזמן בינתיים.

המספר המירבי של התאמות שיש להציג הוא 5. במידה וכבר מוצגות חמש כתבות, ונמצאת כתבה נוספת העונה על ערך החיפוש, יש לדחוף החוצה את הכתבה החמישית, ולהציג את הכתבה החדשה שנמצאה במקום הראשון.

**פרויקט 2: זיהוי טרנדים באתרי חדשות**

תיאור כללי: המערכת תסרוק אתרי חדשות ותחפש בכתבות השונות שלהם את המילים הפופולריות ביותר. המערכת תציג למשתמש שני סוגים של תוצאות:

-רשימת המילים הפופולריות ביותר נכון לרגע זה.

-רשימת מילים שמידת הפופולריות שלהן השתנתה באופן משמעותי בין הרצה להרצה.

פירוט: המערכת תקבל מהמשתמש רשימה של ביטויים שמהם יש להתעלם, ובהתאם לאלה, תסרוק אתרי אינטרנט חדשותיים ותספור את כל שאר המילים. בסיום הסריקה, תציג המערכת סטטיסטיקה על המילים:

-רשימת המילים הפופולריות ביותר בסריקה האחרונה. יוצגו עשר המילים הפופולריות ביותר בסריקה האחרונה, בסדר יורד. כמו כן, תוצג המילה הפופולרית ביותר בכל אתר.

דוגמה לאופן ההצגה:

המילים הפופולריות ביותר:

1.ישראל (26 פעמים)

2.אליפות (22 פעמים)

.....

10. ממשלה (12 פעמים)

פופולריות מילים לפי אתרים:

אתר ONE: המילה אליפות (14 פעמים)

....

את התוכנית יש להריץ כתוכנית גרפית (GUI) שבה יופיעו הפריטים הבאים:

-תיבת טקסט, בה המשתמש יוכל להזין מילים/ביטויים שמהם יש להתעלם בספירה

-כפתור להוספת מילה/ביטוי להתעלמות

-כפתור להתחלת הסריקה

יש לסרוק לפחות שלושה אתרי חדשות באמצעות Web Scrapers.

הסריקה תתבצע בלולאה אינסופית. עם הלחיצה על כפתור התחלת הסריקה, תתחיל הסריקה הראשונה, לאחר שהיא תסתיים, התוכנית תמתין 10 דקות ואז תבצע סריקה נוספת, ותמתין שוב 10 דקות עד לסריקה הבאה.

**פרויקט 3: מערכת להצגת שערי ערך מטבע חוץ**

תיאור כללי: המערכת תציג למשתמש את השווי המשתנה של הכסף שלו, באמצעות סריקת ערכי ההמרה של מטבע חוץ.

פירוט: מסחר במטבע חוץ הוא האפשרות להחליף מטבע מסוים במטבע אחר. לדוגמה, החלפת שקלים בדולרים, לפי שער ההמרה הנקוב של שקל לעומת דולר. שער ההמרה משתנה לאורך שעות פעילות המסחר בכל רגע נתון. זאת אומרת, ששער ההמרה שהיה לפני חמש דקות הוא לא שער ההמרה של עכשיו.

אתרי אינטרנט רבים מציגים את שערי ההמרה של המטבעות השונים.

עליכם לבחור אתר אינטרנט כלשהו שמציג נתוני זמן אמת על שערי מט"ח ולבצע את הפעולות הבאות:

-לחלץ מהאתר את זוגות ההמרה הקיימים בו (לדוגמה, המרה מדולר ליורו – EUR/USD, המרה מפאונד לדולר GBP/USD וכו'). יש להציג על המסך רשימת כפתורים, כשעל כל כפתור שם זוג ההמרה, לדוגמה ERU/USD. יש להציג לכל הפחות חמישה זוגות המרה.

לאחר שהמשתמש ילחץ על אחד הכפתורים, כל הכפתורים ייעלמו ויוצג לו ערך ההמרה של המטבע שבו בחר (ערך זה צריך להתעדכן בכל שנייה), ובנוסף לכך תיבת טקסט שבה יוכל להזין כמה כסף יש לו במטבע 1, ויוצג לו כמה הכסף הזה שווה במטבע השני.

**פרויקט 4: מערכת לזיהוי מדינה לפי כתובת IP**

תיאור כללי: המערכת תאפשר למשתמש להקליד כתובת IP מסוימת, ותציג לו לאיזו מדינה שייכת כתובת ה-IP שהקליד.

פירוט: כל מכשיר המחובר לאינטרנט מבצע את כל ההתקשרויות שלו לרשת באמצעות מזהה הנקרא כתובת IP. תוכלו לבדוק מהי כתובת ה-IP של המכשיר בו אתם משתמשים (מחשב, מובייל), באמצעות חיפוש בגוגל של הביטוי my ip. כתובת ה-IP נקבעת על ידי הנתב/הרשת אליה מחובר המכשיר. כלומר, אם שני מכשירים מחוברים לאותה הרשת, כתובת ה-IP שלהם תהיה זהה.

לכל מדינה המחוברת לאינטרנט, מוקצים מספר טווחי IP. זאת אומרת, שבהינתן כתובת IP, אנו יכולים לדעת לאיזו מדינה כתובת ה-IP הזו שייכת.

האתר <https://lite.ip2location.com/ip-address-ranges-by-country> מרכז את טווחי ה-IP של כל מדינות העולם. המערכת צריכה לסרוק אתר זה, ולשמור את טווחי ה-IP לפי מדינות.

למשתמש יוצג חלון ובו תיבת טקסט. כאשר המשתמש יזין כתובת IP מסוימת, על המערכת לבדוק לאיזו מדינה משתייכת כתובת ה-IP הזו ולהציג את שמה.

**פרויקט 5: מערכת להצגת נתונים מתוך טבלאות ספורט**

תיאור כללי: המערכת תאפשר למשתמש לבחור ליגת ספורט מסוימת ומיקום, ותציג לו את שמה של הקבוצה שנמצאת בליגה זו במיקום הרלוונטי.

פירוט: ישנם הרבה אתרי ספורט המציגים טבלאות של ליגות ספורט (כדורגל, כדורסל וכו'). עליכם לבחור אתר אחד, שעליו תבצעו את פעולת ה-Web Scraping. ראשית, יוצג מסך ובו רשימת הליגות הקיימות (כל ליגה – תוצג ככפתור). יש להציג לפחות חמש ליגות. לאחר שהמשתמש ילחץ על כפתור המייצג ליגה מסוימת, יועבר למסך חדש.

במסך זה, יוצג למשתמש טווח המקומות שיש בליגה, לדוגמה 1-12, והוא יצטרך לבחור מספר בטווח זה וללחוץ על כפתור.

המערכת תיגש אל המיקום הרלוונטי, ותציג את שם הקבוצה ואת מספר הנקודות שלה.

כעבור 10 שניות, המערכת תחזור מחדש אל העמוד הראשי.