**Evghenii Gaisinschii 336551072**

**Lab 1. Summary**

במעבדה זאת נלמדו שיטות וכלים להורדת נתונים מאינטרנט. נבחר להשתמש בשתי ספריות שונות להורדת ציוצים מטוויטר. כמו כן נלמד שיטה לשמירה ואחזור נתונים בבסיס נתונים של **MongoDB**.

ספרייה **tweepy** משתמשת בעזרת כלים שונים להורדת ציוצים מטוויטר דרך פרמטרים שונים שאפשר להגדיר במהלך תכנות. יש מבחר גדול של מטודות שאפשר להשתמש בהם למציאת ציוצים רלוונטיים או לסינון שלהם. כלי הזה דורך פתיחת חשבון מפתח דרך ממשק של טוויטר. נדרש לבקש אישור מצוות של טוויטר שהם מאשרים אחרי כמה שלבים של הזדהות. אחרי קבלת אישור מקבלים כמה טוקנים להפעלת הממשק שלהם.

ספרייה **twint** נותנת גישה להורדת ציוצים מטוויטר דרך מטודות שונות המגדירות סינון של ציוצים לפי פרמטרים שונים כמו זמן עליית הציוץ, שם של משתמש, עיר, כמו הציוצים וכו'. ספרייה די פשוטה שלא דורשת ממשקים חיצוניים ופתיחת חשבונות נוספים.

בנוסף הורדנו **MongoDB Compass** לצורך שמירה ואחזור נתונים שהורדנו. כלי נותן ממש ידידותי להבנת ארכיטקטורת מבנה נתונים ובדיקה האם נתונים נשמרים כמו שצריך. שימוש נוח בהפעלת שרת וייצוא וייבוא נתונים ממאגרים נשמרים.

בעזר כלים הנ"ל נבנה מערכת המחפשת ומורידה ציוצים מטוויטר לפי פרמטרים שונים. המערכת יכולה גם להוריד ציוצים וגם לחפש ציוצים שכבר הורדנו ושמרנו בשרת.

נדגים כל הפונקציונליות של מערכת:

בתמונה למטה אפשר לראות ממש משתמש של מערכת. יש לנו אופציה לחפש לפי פרמטרים ספציפיים בספריית tweepy ו-twint. בצד שמאל אפשר לחפש ולהוריד ציוצים לפי שם הנושא, כמות הציוצים שנרצה להוריד וגם שפה בה ציוצים נכתבים. בצד ימין אפשר לחפש ולהוריד ציוצים לפי עיר והגבלת הציוצים.

Diagram, waterfall chart

Description automatically generated

בנוסף מתחת לכל חיפוש יש אופציה לראות במאגרי נתונים שחיפשנו ושמרנו לפי כל ספרייה מה הם הציוצים שקיבלו הכי הרבה לייקים, הכי הרבה רטוויטים, מה הציוצים שהכי חוזרים על עצמם ולאיזה משתמש יש הכי הרבה עוקבים. כל נתון זה מוצג במסך עם פירוט של כמות לייקים, רטוויטים ,עוקבים וציוצים עצמם.

תמונות הבאות אפשר לראות את כל הפונקציונליות של מערכת וחיפוש נתונים עבור כל פרמטר שצוין לפני.

Diagram

Description automatically generated

Diagram

Description automatically generated

Diagram

Description automatically generated

Diagram

Description automatically generated

Diagram

Description automatically generated

Diagram

Description automatically generated