**תרגיל בית 1:**

**ישראל שושן 318740834**

**דוד אסולין 205908569**

**עמית פרץ 209054303**

**יובל הילאי 318609237**

[**GitHub**](https://github.com/AmitPerets/Index_Invincibles_HW_1)

**שאלה 1 – סעיף א :**

1. **4a. Term-Document Incidence Matrix:**

נתונים:

* **Doc 1:** breakthrough drugs for schizophrenia
* **Doc 2:** new schizophrenia drugs
* **Doc 3:** new approach for treatment of schizophrenia
* **Doc 4:** new hopes for schizophrenia patients

ראשית, נרשום את כל המונחים הייחודיים באוסף המסמכים:

* breakthrough
* drugs
* for
* schizophrenia
* new
* approach
* treatment
* hopes
* patients

נבנה מטריצת שכיחות מונח במסמך.

שורות מייצגות מונחים ועמודות מייצגות מסמכים. ערך 1 מציין את נוכחותו של המונח במסמך, ו-0 מציין את היעדרו.

A table with numbers and letters

Description automatically generated

**4b. Inverted Index Representation:**

נבנה את ייצוג האינדקס ההפוך, כל מונח ישויך לרשימה של מזהי מסמכים שבהם הוא מופיע.

* breakthrough: [1]
* drugs: [1, 2]
* for: [1, 4]
* schizophrenia: [1, 2, 3, 4]
* new: [2, 3, 4]
* approach: [3]
* treatment: [3]
* hopes: [4]
* patients: [4]

תוצאות שהוחזרו עבור שאילתות:

עבור אוסף המסמכים המוצג לעיל, נמצא את התוצאות שהוחזרו עבור השאילתות הנתונות:

א. שאילתה: schizophrenia and drugs

השאילתה schizophrenia and drugs דורשת מסמכים המכילים גם את המונחים " schizophrenia " וגם " drugs". מהאינדקס ההפוך, יש לנו:

* schizophrenia: [1, 2, 3, 4]
* drugs: [1, 2]

החיתוך של רשימות אלה נותן לנו:

[1, 2]

אז המסמכים שמוחזרים הם: Doc1 וDoc2

ב. שאילתה: בעד ולא (סמים או גישה)

השאילתה עבור ולא (סמים או גישה) דורשת מסמכים המכילים את המונח "עבור" אך אינם מכילים את המונחים "סמים" או "גישה". מהאינדקס ההפוך, יש לנו:

* for: [1, 4]
* drugs: [1, 2]
* approach: [3]

ראשית, אנו מוצאים את האיחוד של "סמים" ו"גישה":

* drugs or approach: [1, 2, 3]

לאחר מכן, אנו שוללים רשימה זו כדי למצוא מסמכים שאינם מכילים "סמים" או "גישה":

* not (drugs or approach): [4] (since only Doc 4 is not in [1, 2, 3])

לבסוף, אנו עושים את החיתוך בין התוצאה הזו עם רשימת המסמכים המכילה "עבור":

* for: [1, 4]
* not (drugs or approach): [4]

החיתוך נותן לנו

[4]

אז המסמך המוחזר הוא **Doc 4**

**שאלה 1 – סעיף ב :**

שלבי חישוב :

* 1. סה''כ Terms ב Doc1 : 27+3+14 = 44
  2. סה''כ Terms ב Doc2 : 4 + 33 + 33 = 70
  3. סה''כ Terms ב Doc3: 24+29+17 = 70

**TF :**

|  |  |  |  |
| --- | --- | --- | --- |
| Doc3 | Doc2 | Doc1 |  |
|  |  |  | car |
| 0 |  |  | auto |
|  |  | 0 | insurance |
|  |  |  | best |

TF-IDF:

|  |  |  |  |
| --- | --- | --- | --- |
| Doc3 | Doc2 | Doc1 |  |
|  |  |  | car |
| 0 |  |  | auto |
|  |  | 0 | insurance |
|  |  |  | best |

3. תחום העיסוק המרכזי של אתר EA SPORTS הוא פיתוח והפצת משחקי וידאו בתחום הספורט, במיוחד משחקי כדורגל. האתר מספק מידע רחב על המשחקים השונים של החברה, כולל EA SPORTS FC™ 24, FIFA, Madden NFL, NHL, UFC, ו-Apex Legends. המשתמשים יכולים למצוא מידע על עדכונים חדשים, תכונות משחק, מצבי משחק שונים, חדשות על תחרויות, פרטים על שחקנים וקבוצות, הוראות קנייה, ומבצעים מיוחדים. האתר גם מציע גישה לפורומים, עדכוני דוא"ל ופרטי תמיכה טכנית.

EA SPORTS: <www.ea.com>

שאילתא: מידע על האירועים הקהילתיים המתוכננים על ידי EA SPORTS

מידע נדרש: פרטי אירועים קהילתיים כמו תאריכים, מיקומים, פעילויות מתוכננות, ואיך להשתתף בהם.

שאילתא: השפעת העדכונים השונים על ביצועי המשחק

מידע נדרש: השוואה בין הגרסאות השונות של המשחק, כולל הבדלים בביצועים, תכונות חדשות, ותיקונים שבוצעו.

שאילתא: נתונים סטטיסטיים על השחקנים הטובים ביותר במשחקי EA SPORTS

מידע נדרש: רשימות וסטטיסטיקות של השחקנים המובילים בכל משחק, כולל נתוני ביצועים ותחזיות לשיפור עתידי.