**אחזור מידע – תרגיל בית 1 – אביב התשפ"ד**

**מגישים:**

**מור שמואל 315040980**

**מעיין אביטן 313550907**

**אופק בן אברהם 313310559**

**רותם פורת 206607731**

שאלה 1: מודלים לאחזור מידע

שאלה 1.4a.a -

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **D1** | **D2** | **D3** | **D4** |
| **Breakthrough** | 1 | 0 | 0 | 0 |
| **Drugs** | 1 | 1 | 0 | 0 |
| **For** | 1 | 0 | 1 | 1 |
| **Schizophrenia** | 1 | 1 | 1 | 1 |
| **New** | 0 | 1 | 1 | 1 |
| **Approach** | 0 | 0 | 1 | 0 |
| **Treatment** | 0 | 0 | 1 | 0 |
| **Of** | 0 | 0 | 1 | 0 |
| **Hopes** | 0 | 0 | 0 | 1 |
| **Patients** | 0 | 0 | 0 | 1 |

שאלה 1.4a.b -

Breakthrough -> [1]

Drugs -> [1,2]

For -> [1,3,4]

Schizophrenia -> [1,2,3,4]

New -> [2,3,4]

Approach -> [3]

Treatment -> [3]

Of -> [3]

Hopes -> [4]

Patients -> [4]

שאלה 1.4b.a -

schizophrenia and drugs - [1,1,1,1] AND [1,1,0,0] = [1,1,0,0] = D1, D2

שאלה 1.4b.b -

for and not (drugs or approach) - [1,0,1,1] AND NOT ([1,1,0,0] AND [0,0,1,0]) = [1,0,1,1] AND NOT ([0,0,0,0]) = [1,0,1,1] AND [1,1,1,1] = [1,0,1,1] = D1, D3, D4

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **D3 tf-idf** | **D2 tf-idf** | **D1 tf-idf** | **idf** | **df** | **D3** | **D2** | **D1** |  |
| 39.60 | 6.6 | 44.55 | 1.65 | 18165 | 24 | 4 | 27 | **car** |
| 0 | 68.64 | 6.24 | 2.08 | 6723 | 0 | 33 | 3 | **auto** |
| 46.98 | 53.46 | 0 | 1.62 | 19241 | 29 | 33 | 0 | **insurance** |
| 25.5 | 0 | 21 | 1.5 | 25235 | 17 | 0 | 14 | **best** |

הסבר: העמודות השמאליות של Doc1 , Doc2, Doc3 הם לפי הנתונים. שאר העמודות הם לפי הנוסחאות שנלמדו בהרצאה.

שאלה 2: בניית אתר HTML

**קישור לGIT:**

[**https://github.com/avtn96/Information\_Retrieval**](https://github.com/avtn96/Information_Retrieval)

שאלה 3: קדם פרויקט –חקר אתר

האתר שבחרנו הוא IMDb:

<https://www.imdb.com/>

1. האתר IMDb הוא אתר המתמקד בסרטים, טלוויזיה, סדרות וכוכבים. האתר מספק מידע על סרטים ותכני תרבות, כולל נתונים כגון דירוגים, ביקורות, תקצירים, מידע על צוות הפקה ושחקנים, פרסומות ועוד. המשתמשים יכולים לחפש סרטים וסדרות לפי שם, שנת יציאה, ז'אנר, דירוגים ולקבל מידע מקיף על יצירות תרבות משונות.
2. שלוש שאילתות מעניינות שהיינו רוצים לקבל עליהן תשובה באתר והאתר אינו עונה עליהן כעת:

* Films that won the Academy Award for Best Non-Comedy Film since 2000

**פרטי המידע הנדרשים**: רשימת סרטים, שנת הזכייה, קטגוריה המתארת את סוג הפרס.

* Stars who starred in more than four films that won the Oscar for Best Picture

**פרטי המידע הנדרשים**: שמות הכוכבים, רשימת הסרטים שבהם הם שיחקו.

* Highest rated TV series on IMDb that started airing more than 20 years ago

**פרטי המידע הנדרשים**: רשימת הסדרות, הדירוג הגבוה ביותר, שנת התחילת השידור.