דינה קקיאשוילי 318296936

נטלי נניקשוילי 208744094

1. א) פתרון שאלות 4a, 4b מתרגול 2:

4a

תחילה נרשום עבור כל מסמך את המילים שיש בו ונייצר מכך טבלה:

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Doc 1 | Doc 2 | Doc 3 | Doc 4 |
| 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 |

כעת נייצג את התוצאה שקיבלנו כפי שניתן בדוגמה :

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 4b:

**schizophrenia AND drugs** :

schizophrenia: 1 1 1 1

drugs: 1 1 0 0

1 & 1 = 1

1 & 1 = 1

1 & 0 = 0

1 & 0 = 0

ומכאן שהמסמכים שמוחזרים הם מסמך 1 ומסמך 2.

**for AND NOT (drugs OR approach)**

תחילה נחשב עבור drugs OR approach :

drugs: 1 1 0 0

approach: 0 0 1 0

for: 1 0 1 1

1 || 0 = 1

1 || 0 = 1

0 || 1 = 1

0|| 0 = 0

ומכך נסיק שהNOT של מה שיצא זה 1 0 0 0

כעת נחשב את הביטוי for AND NOT (drugs OR approach) :

1 & 0 = 0

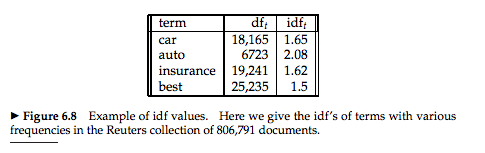
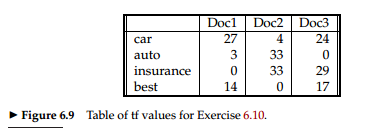
0 & 0 = 0

1 & 0 = 0

1 & 1 = 1

ומכאן שהמסמך המבוקש הוא מסמך 4.

1. ב) פתרון tf-idf



|  |  |  |
| --- | --- | --- |
| idf | df | Terms |
| 1.65 | 18165 | car |
| 2.08 | 6723 | Auto |
| 1.62 | 19421 | Insurance |
| 1.5 | 25235 | best |

|  |  |  |  |
| --- | --- | --- | --- |
| D3 | D2 | D1 | Terms |
| 39.6 | 6.6 | 44.55 | car |
| 0 | 68.64 | 6.24 | Auto |
| 46.98 | 53.46 | 0 | Insurance |
| 25.5 | 0 | 21 | best |

|  |  |  |  |
| --- | --- | --- | --- |
| Tf D3 | Tf D2 | Tf D1 | Terms |
| 24 | 4 | 27 | car |
| 0 | 33 | 3 | Auto |
| 29 | 33 | 0 | Insurance |
| 17 | 0 | 14 | best |

|  |  |
| --- | --- |
| Vector length | doc |
| 49.6451 | D1 |
| 87.2524 | D2 |
| 66.5247 | D3 |

3) פתרון השאלה מהתרגול חקר אתר:

**האתר שנבחר:** boohoo

https://www.boohoo.com/

**תחום העיסוק של האתר**: Boohoo היא פלטפורמת מסחר אלקטרוני המתמחה באופנה עדכנית במחירים נגישים, ומציעה מגוון רחב של מוצרים כולל בגדים, נעליים, אביזרים ומוצרי יופי. האתר פונה לצרכנים צעירים ואופנתיים ומציע אוסף אוצרות של פריטים טרנדיים ממגוון מותגים.

**שאילתות מעניינות:**

שאילתה: best seller

פרטי מידע נדרשים: כל המוצרים שמקבלים את התגית best seller

**שאילתה: מה כמות המוצרים במלאי מחולצה מסוימת?**

פרטי מידע נדרשים: כמות המוצרים שנמצאים במלאי, מספר מזהה של החולצה

**שאילתה: המותג הכי נמכר**

פרטי מידע נדרשים: מותגים פופולריים ודפוסי מכירה לפי חודשים.