# Information Retrieval

## Search Engine Bandits

#### Sara Asaad – 207674318

#### Ward Zidani – 318531944

#### Safwan Halabi – 213784127

[GitHub Link](https://github.com/Safwan-Halabi/Information-Retrieval.git): - <https://github.com/Safwan-Halabi/Information-Retrieval.git>

**Question 1.1:**

**Exercise 4a from Classroom Practice 2:**

Consider these documents:

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

a. Draw the term-document incidence matrix for this document collection.

b. Draw the inverted index representation for this collection, as in Figure 1.3.

**Answer:**

4.a.a:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Term | 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 |

4.a.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 | - | - | - |

**Exercise 4b from Classroom Practice 2:**

For the document collection shown in Exercise 4a, what are the returned results for these queries?

a. “schizophrenia” and “drugs”.

b. “for” and **not** (“drugs” or “approach”).

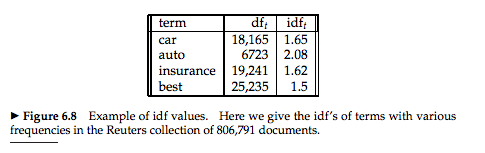
**Answer:**

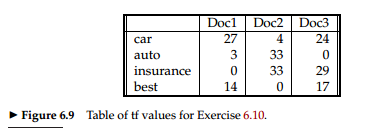
4.b.a:

The returned result is {D1, D2}.

4.b.b:

The returned result is {D4}.

**Question 1.2:**



Calculate the TF-IDF value for each term and for each document:

TF = Term Frequency.

IDF = Inverse Document Frequency.

**Answer:**

“Car”:

TF-IDF(“Car”, Doc1) =

TF-IDF(“Car”, Doc2) =

TF-IDF(“Car”, Doc3) =

“Auto”:

TF-IDF(“Auto”, Doc1) =

TF-IDF(“Auto”, Doc2) =

TF-IDF(“Auto”, Doc3) =

“Insurance”:

TF-IDF(“Insurance”, Doc1) =

TF-IDF(“Insurance”, Doc2) =

TF-IDF(“Insurance”, Doc3) =

“Best”:

TF-IDF(“Best”, Doc1) =

TF-IDF(“Best”, Doc2) =

TF-IDF(“Best”, Doc3) =

**Question 3:**

1. מהו תחום העיסוק המרכזי של האתר? מהו המידע הזמין למשתמשי האתר? ענו בפסקה אחת. צרפו את הקישור לאתר.

**תשובה:**

האתר [**LinkedIn**](https://www.linkedin.com/in/ward-zidani/) הוא רשת חברתית מקצועית המשמשת למטרות עבודה וקריירה. התחום העיסוק המרכזי של האתר הוא חיבור מקצועי בין משתמשים לצורך בניית רשתות קשרים מקצועיים, חיפוש עבודה, גיוס עובדים, ושיתוף מידע מקצועי ותוכן מקורי. למשתמשי האתר זמין מידע רב הכולל פרופילים מקצועיים של אנשי מקצוע ועסקים, פרסומים ומאמרים בתחום הקריירה והעבודה, משרות פנויות, המלצות וקשרים בין משתמשים, קבוצות דיון בנושאים מקצועיים, והזדמנויות לימודיות והכשרה מקצועית. ניתן גם ליצור ולנהל דפים עסקיים לפרסם תוכן לקידום מותגים וחברות. לפרטים נוספים, בקרו באתר: (<https://www.linkedin.com>).

2.     רשמו שלוש שאילתות מעניינות שהייתם רוצים לקבל עליהן תשובה באתר, והאתר אינו **עונה עליהן** כעת. עבור כל שאילתא כזו, ציינו מהם פרטי המידע הנדרשים לצורך מענה על השאילתא.

**תשובה:**

1. הצגת התחומים והטכנולוגיות המבוקשות ביותר בתעשייה (לפי משרות חדשות שנפתחו).
2. הצגת סטטיסטיקות של חיפוש עבודה לפי רמות ניסיון ותחומי התמחות (חיפוש לפי רמת ניסיון דרוש).
3. להציע תחומים חסרים לפי משרה\משרות - מותאם אישית לפרופיל המשתמש (הצעת קורסים\הכשרות למשתמש).