**1. 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

2. ייצוג אינדקס הפוך

לאחר מכן, 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]

3. 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]

**Doc 1** ו**Doc 2** אז המסמכים שמוחזרים הם :

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**