**תרגיל בית 2 – אחזור מידע**

שמות המגישים: ברק סגל – 204336440, מאור ברדוגו –308077411, נתנאל חג'ג' - 200388676



|  |  |
| --- | --- |
| Term | Pages |
| A | 1 |
| Also | 4 |
| And | 1,3 |
| Can | 3 |
| Course | 2 |
| Dog | 1,3 |
| From | 4 |
| He | 3 |
| Help | 4 |
| Information | 2 |
| Listen | 1 |
| Lot | 1 |
| Love | 1,3 |
| Me | 4 |
| Music | 1 |
| My | 1 |
| Often | 1,4 |
| Pick | 4 |
| Retrieval | 2 |
| Road | 4 |
| roll | 1,3 |
| Stone | 1,3,4 |
| The | 1,3,4 |
| They | 4 |
| Throw | 3 |
| To | 1 |
| Up | 4 |



|  |  |
| --- | --- |
| Term | Pages |
| A | 1 |
| Also | 4 |
| And | 1,3 |
| Can | 3 |
| Course | 2 |
| Dog | 3 |
| Dogs | 1 |
| From | 4 |
| He | 3 |
| Help | 4 |
| Information | 2 |
| Listen | 1 |
| Lot | 1 |
| Love | 1 |
| Loves | 3 |
| Me | 4 |
| Music | 1 |
| My | 1 |
| Often | 1,4, |
| Pick | 4 |
| Retrieval | 2 |
| Road | 4 |
| roll | 3 |
| Rolling | 1,3 |
| Stones | 1,3,4 |
| The | 1,3,4 |
| They | 4 |
| Throwing | 3 |
| To | 1 |
| Up | 4 |
|  |  |

|  |  |  |
| --- | --- | --- |
| Rank | Term | cnt |
| 1 | the | 3 |
| 2 | stones | 3 |
| 3 | and | 2 |
| 4 | often | 2 |
| 5 | rolling | 2 |
| 6 | my | 1 |
| 7 | dog | 1 |
| 8 | love | 1 |
| 9 | music | 1 |
| 10 | a | 1 |
| 11 | lot | 1 |
| 12 | listen | 1 |
| 13 | to | 1 |
| 14 | roll | 1 |
| 15 | dogs | 1 |
| 16 | loves | 1 |
| 17 | Information | 1 |
| 18 | Retrieval | 1 |
| 19 | course | 1 |
| 20 | can | 1 |
| 21 | he | 1 |
| 22 | throwing | 1 |
| 23 | They | 1 |
| 24 | also | 1 |
| 25 | help | 1 |
| 26 | me | 1 |
| 27 | pick | 1 |
| 28 | up | 1 |
| 29 | from | 1 |
| 30 | road | 1 |

לפי הגרף ניתן לראות שמרבית המילים מופיעות אך ורק פעם אחת ולכן לדעתנו המילים מתנהגות לפי חוק zipf.



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| weights=tf-idf | | | Idf | tf | | |  |
| Doc3 | Doc2 | Doc1 | Doc3 | Doc2 | Doc1 |
|  |  |  | 1.65 |  |  |  | car |
| 0 |  |  | 2.08 | 0 |  |  | auto |
|  |  | 0 | 1.62 |  |  | 0 | Insurance |
|  | 0 |  | 1.5 |  | 0 |  | Best |

החישוב של tf-idf מתבצע על ידי כפל של כל ערך שבתא ה- tf בערך שבתא ה-idf.

|RET|=18, |REL|=20, |RET∩REL|=8,

1. מנוע 1:  
     
   רלוונטים מסומנים בירוק.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Precision | 1 |  |  |  |  |  |  |  |  |  |
| Recall |  |  |  |  |  |  |  |  |  |  |

מנוע 2:  
  
רלוונטים מסומנים בירוק.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Precision | 0 |  |  |  |  |  |  |  |  |  |
| Recall | 0 |  |  |  |  |  |  |  |  |  |

**שאלה 2:**

חלק א:

1. black
2. office
3. insprion

חלק ב: קדם פרויקט –בניית זחלן

זחלן ממומש מצורף בקובץ פייתון.

1.

שאילתא: “black”

|  |  |  |
| --- | --- | --- |
| **Word** | **Inverted index** | **count** |
| Processor | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| inch | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| DDR4 | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| 1TB | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| battery | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| laptop | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| more | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| details | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| processor | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 20 | 19 |
| Operating | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 19 |
| with | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 19 |
| Display | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 19 |
| Storage | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 19 |
| HDD | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 19 |
| Design | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 19 |

שאילתא: "office"

|  |  |  |
| --- | --- | --- |
| **Word** | **Inverted index** | **count** |
| Windows | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| 10 | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| RAM | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 18-> 19-> 20 | 20 |
| processor | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 15-> 16-> 18-> 19-> 20 | 19 |
| inch | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 15-> 16-> 18-> 19-> 20 | 19 |
| Graphics | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 15-> 16-> 18-> 19-> 20 | 19 |
| more | 0-> 1-> 2-> 3-> 5-> 6-> 7-> 8-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 19 |
| details | 0-> 1-> 2-> 3-> 5-> 6-> 7-> 8-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 19 |
| drive | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 15-> 16-> 18-> 19-> 20 | 19 |
| Intel | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 16-> 17-> 18-> 19-> 20 | 18 |
| DDR4 | 0-> 1-> 2-> 4-> 5-> 6-> 7-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 20 | 18 |
| laptop | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 15-> 16-> 18-> 20 | 18 |
| Home | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 17-> 18-> 19 | 17 |
| Display | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 17-> 18 | 16 |
| and | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 13-> 14-> 17-> 18 | 16 |

שאילתא: "“ inspiron

|  |  |  |
| --- | --- | --- |
| **Word** | **Inverted index** | **count** |
| ram | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| more | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 21 |
| processor | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 20 | 20 |
| details | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 20 |
| see | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 20 |
| drive | 0-> 1-> 2-> 3-> 4-> 5-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 17-> 18-> 19-> 20 | 19 |
| operating | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 17-> 18 | 18 |
| system | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 17-> 18 | 18 |
| windows | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 17-> 18-> 19 | 18 |
| 10 | 0-> 1-> 2-> 3-> 4-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 17-> 18-> 19 | 18 |
| graphics | 0-> 1-> 2-> 3-> 4-> 5-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 17-> 18-> 20 | 18 |
| inch | 0-> 1-> 2-> 3-> 4-> 5-> 7-> 8-> 10-> 11-> 12-> 13-> 14-> 15-> 17-> 18-> 20 | 17 |
| product | 0-> 2-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 18-> 19-> 20 | 17 |
| laptop | 0-> 1-> 2-> 3-> 4-> 5-> 7-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 16-> 19 | 16 |
| › | 0-> 2-> 5-> 6-> 8-> 9-> 10-> 11-> 12-> 13-> 14-> 15-> 16-> 17-> 19-> 20 | 16 |

2.

|  |  |  |  |
| --- | --- | --- | --- |
| Tf.idf | tf | Query=”office” | Docs |
|  | 1+log(1)=1 | 1 | doc0 |
|  | 1+log(1)=1 | 1 | doc1 |
|  | 1+log(1)=1 | 1 | doc2 |
|  | 1+log(1)=1 | 1 | doc3 |
| 0 | 0 | 0 | doc4 |
|  | 1+log(1)=1 | 1 | doc5 |
|  | 1+log(1)=1 | 1 | doc6 |
|  | 1+log(1)=1 | 1 | doc7 |
|  | 1+log(1)=1 | 1 | doc8 |
| 0 | 0 | 0 | doc9 |
|  | 1+log(1)=1 | 1 | doc10 |
|  | 1+log(1)=1 | 1 | doc11 |
| 0 | 0 | 0 | doc12 |
|  | 1+log(1)=1 | 1 | doc13 |
| 1+log(1)=1 | 1+log(1)=1 | 1 | doc14 |
| 0 | 0 | 0 | doc15 |
| 0 | 0 | 0 | doc16 |
|  | 1+log(1)=1 | 1 | doc17 |
|  | 1+log(1)=1 | 1 | doc18 |
| 0 | 0 | 0 | doc19 |

3.

השאילתה : “black laptops” באתר אמזון.

Recall of Netanel: , Recall of Barak:

Relevant

Idan

Hen

Relevant

Idan

Hen

Percission of Netanel: , Percission of Barak:

Relevant

Idan

Hen

Relevant

Idan

Hen

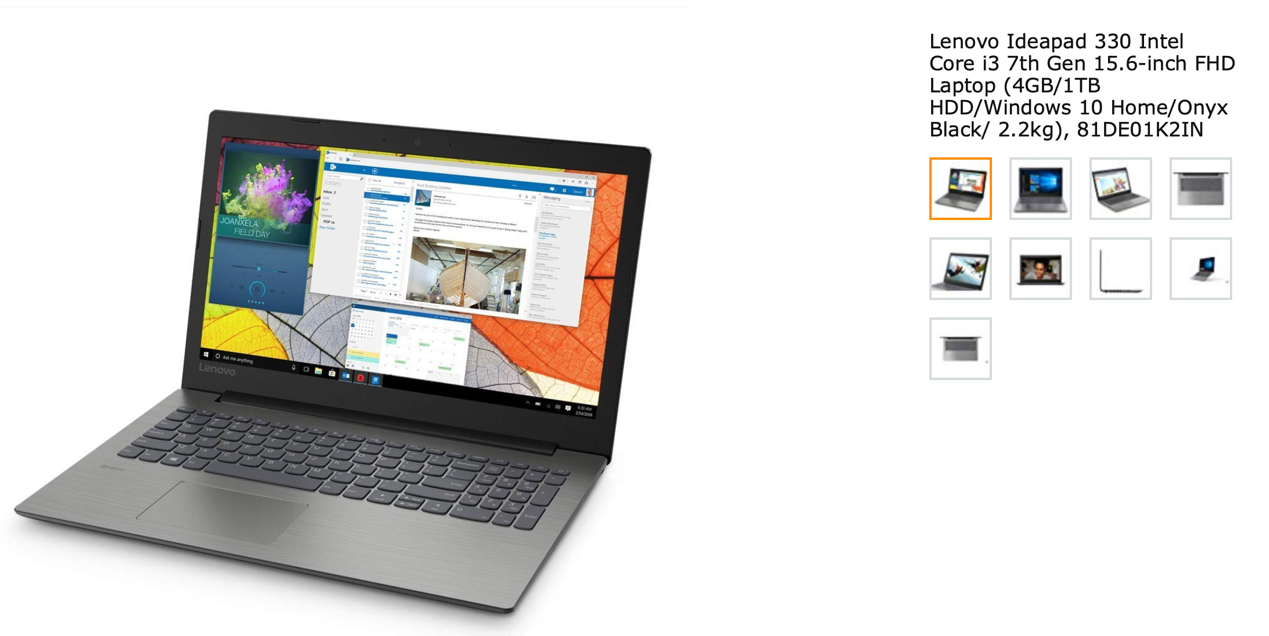
הסבר: בגלל אופן החיפוש בזחלן , שמוגדר לחפש את כל המוצרים בסקשן “laptops” ע״פ URL מתאים שסיפקנו לו ומתוך המוצרים שהזחלן רואה בסקשן הזה , הוא מחזיר רק את התוצאות שהמילה black מופיעה בתיאור המוצר, זה גרם לאחוזי דיוק גבוהים מאוד כי צבע כמעט תמיד מוזכר בתיאור המוצר רק בהקשר של צבע צבע המוצר.



















1. 