אחזור מידע מטלה מספר 2

מגישים: אריק זגדון - 204459093

דביר בן סימון - 308315035

שנהב חזי - 313594764

חלק א

פיתרון שאלה 1:

|  |  |  |
| --- | --- | --- |
| Q | Engine 1 | Engine 2 |
| 1 | Recall = 0.17  Precision = 1.0 | Recall = 0.0  Precision = 0.0 |
| 2 | Recall =0.17  Precision = 0.5 | Recall = 0.2  Precision = 0.5 |
| 3 | Recall = 0.34  Precision = 0.67 | Recall = 0.4  Precision = 0.67 |
| 4 | Recall = 0.5  Precision = 0.75 | Recall = 0.6  Precision = 0.75 |
| 5 | Recall = 0.5  Precision = 0.6 | Recall = 0.6  Precision = 0.6 |
| 6 | Recall = 0.5  Precision = 0.5 | Recall = 0.8  Precision = 0.67 |
| 7 | Recall = 0.67  Precision = 0.57 | Recall = 1.0  Precision = 0.71 |
| 8 | Recall = 0.67  Precision = 0.5 | Recall = 1.0  Precision = 0.62 |
| 9 | Recall = 0.84  Precision = 0.55 | Recall = 1.0  Precision = 0.55 |
| 10 | Recall = 1.0  Precision = 0.6 | Recall = 1.0  Precision = 0.5 |

פיתרון שאלה 2:

פיתרון שאלה 3:

א+ב.

|  |  |
| --- | --- |
| Dictionary | Posting List |
| course | 2 |
| dog | 1,3 |
| help | 4 |
| information | 2 |
| listen | 1 |
| love | 1,3 |
| music | 1 |
| pick | 4 |
| road | 4 |
| retrieval | 2 |
| roll | 1,3 |
| stone | 1,3,4 |
| throw | 3 |

ג.

|  |  |
| --- | --- |
| Rank | Frequency |
| 1 | 3 |
| 2 | 2 |
| 3 | 2 |
| 4 | 1 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |
| 11 | 1 |
| 12 | 1 |

ניתן לראות לפי הגרף כי אכן ההתנהגות זהה לפי חוק ZIPY.

שאלה 2:

חלק א:

המשתמש בוחר כמות מניות(5/10/20/25) מתוך המניות הכי פעילות לפי האתר והנתונים שיוחזרו עבור כל מניה הם:

* שם המניה
* סקטור
* תעשייה
* מחיר מניה נוכחי
* מחיר מניה צפוי עפ"י אנליסטים
* רווח
* רווח באחוזים
* אומדן רווחים
* אומדן רווחים באחוזים
* רווח למניה
* דיבידנדים עתידיים

תגיות שנעשה בהן שימוש: <h1>, <td>, <p>, <span>, <table>, <tr>

השאילתות:

1. X המניות בעלות חיזוי רווח מחיר מניה הגבוהה ביותר באחוזים מהמניות הכי פעילות בשוק.

הסבר: ביצוע מעבר על כל המניות הכי פעילות בשוק עפ"י Yahoo Finance, חישוב הפרש בין מחיר נוכחי לבין חיזוי אנליסטים עבור מחיר המניה בסוף השנה, חישוב היחס בין המחיר הנוכחי להפרש והמרה לאחוזים.

1. X המניות בעלות חיזוי רווח למניה(EPS – Earning Per Share) הגבוהה ביותר באחוזים מהמניות הכי פעילות בשוק.

הסבר: ביצוע מעבר על כל המניות הכי פעילות בשוק עפ"י Yahoo Finance, חישוב הפרש בין מחיר למניה נוכחי(EPS) לבין חיזוי אנליסטים עבור מחיר למניה(EPS) בסוף השנה, חישוב היחס בין המחיר הנוכחי להפרש והמרה לאחוזים.

1. X החברות אשר מעניקות דיבידנד לבעלי המניות שלהן באחוזים מהמניות הכי פעילות בשוק.

הסבר: ביצוע מעבר על כל החברות בעלות המניות הכי פעילות בשוק עפ"י Yahoo Finance, יוחזרו כל החברות שקים עבורם מידע על דיבידנד.

חלק ב:

1.

הזחלן שלנו עובר על כל אחת מהמניות שמופיעות באותה הזמן במניות הכי פעילות.

ולכן התוצאות של המילים הכי נפוצות יהיו זהות בכל אחד מדפי המנייה שחזרו מהזחלן.

ולכן בתוצאות שנציג כעת המילים הנפוצות נמצאות בכל הדפים.

רשימת המילים הנפוצות ביותר:

Estimate, Day, Stocks, Market, Finance, Inc, Analysis, Price, Change, Qtr, Year, Futures, Buy, Sell, News

רשימת הinverted-index :

**Estimate** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Stocks** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Market** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Sell** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Finance** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Qtr** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Inc** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Day** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Analysis** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Price** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Change** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Futures** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Buy** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**News** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

**Year** -> APPL,WBD,AMZN,TSLA,F,MSFT,LUMN,INTC,T,CMCSA,META,CSCO,PFE,VZ,AMD,BAC,RIVN,GOOGL,XOM,BKR

2:

השאילתה שבחרנו לעשות עלייה tf-id היא שאילתא 3 :

X החברות אשר מעניקות דיבידנד לבעלי המניות שלהן באחוזים מהמניות הכי פעילות בשוק.

הterms שעילהם נציג : Companies, Dividend, stocks, market :

מהשאילתה חזרו לנו 5 מסמכים (דפי מניות ) המניות שחזרו הן : APPL, F,MSFT,INTC,T

N=20 כל המניות שבדקנו.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| T | INTC | MSTF | F | APPL | TF-ID |
| 0.602 | 0.692 | 0.889 | 0.602 | 0.889 | Companies |
| 1.229 | 0.964 | 0.783 | 0.783 | 0.964 | Dividend |
| 1.229 | 1.292 | 1.357 | 1.457 | 1.251 | Stocks |
| 1.327 | 1.110 | 1.310 | 1.204 | 1.251 | market |

3: נענה על שאלה זאת עבור השאילתא השלישית כמו בשאלה קודמת :

הצגנו לשני חברים שעוסקים בתחום הכלכלה את 10 המניות הראשונות שחזרו מהשאילתא ושאלנו אותם אילו מ10 המניות הללו רלוונטיות עבורם לאחר מכן אנו מבצעים את החישוב עבור recall

Precition

|  |  |  |
| --- | --- | --- |
|  | Relevant | Nonrelevant |
| Retrieved | 6(tp) | 4(fp) |
| Not Retrieved | 6(fn) | 3(tn) |