

1. 5% ההשפעה של חוק zipf באה לידי ביטוי באורך רשימות postings של terms על פי הפירוט הבא (רשימות posting הן הרשימות הכוללות אינפורמציה על מופעים של Terms במסמכים):

אין קשר בין הדברים

רשימות של מעט מילים תהיינה מאוד ארוכות ושל הרבה מילים תהיינה מאוד קצרות
רשימות של הרבה מילים תהיינה מאוד ארוכות ושל מעט מילים תהיינה מאוד קצרות
החוק יבוא לידי ביטוי רק אם יופעל תהליך stemming על המסמכים
החוק יבוא לידי ביטוי רק אם יופעל תהליך של הסרת stopwords

ב+ד נכונים

ב+ה נכונים

ג+ה נכונים

2. 4% מדדי ההערכה הבאים מתאימים להערכת מנועי חיפוש אינטרנטיים:

DCG , F-measure, precision@k, R-precision
MAP , interpolated average precision , precision@k , DCG
Reciprocal Rank , precision@k , MAP , DCG
MAP , precision@10 , Fallout , NDCG

כל המדדים שהוזכרו בהרצאות יכולים להתאים גם למנועי חיפוש אינטרנטיים
אף אחד מהמדדים שהוזכרו בהרצאות לא יכול להתאים בלי שינויים בהגדרות

3. 4% הפרמטר d בנוסחת Pagerank

מאפשר למדל באופן נכון את התנהגות ה Random surfer (הגולש האקראי) ולפתור את בעיית ה spider trap

מאפשר לחשב pagerank באופן מהיר

מאפשר לנרמל את תוצאת ה Pagerank

הוא הווקטור העצמי של מטריצת המעברים

מאפשר למשתמש לעבור לדף באופן אקראי

א+ה נכונים

4. 4% עדכון של האינדקס מיד עם הגעה של מסמך חדש מה crawler
מאט את פעולת האינדוקס
משפר בהכרח את ה precision של שאילתות רלוונטיות
משפר בהכרח את ה recall של שאילתות רלוונטיות
מעלה את העדכניות של האינדקס אך מסבך את הניהול שלו
משפר את המהירות של ביצוע השאילתא.

5. 5% נתון מאגר עם 4 המסמכים הבאים:

D_1 : alpha bravo charlie delta echo foxtrot golf

D_2 : golf golf golf delta alpha

D_3 : bravo charlie bravo echo foxtrot bravo

D_4 : foxtrot alpha alpha golf golf delta

לשאילתא alpha bravo

מנוע שמבוסס על vector space יחזיר את כל המסמכים ואת מסמך D_1 במקום הראשון
מנוע שמתבסס על vector space יחזיר את כל המסמכים ואת מסמך D_3 במקום הראשון
מנוע שמתבסס על מודל בוליאני טהור יחזיר את מסמך d_1
מנוע שמתבסס על מנוע בוליאני טהור לא יחזיר את מסמך d_2
ג+ד נכונים
ב+ג נכונים

סמן נכון או לא נכון:

6. 4% כדי לאפשר מענה לשאילתות של ביטויים האינדקס חייב להיות במבנה של
biword index .

7. 4% Relevance feedback יכול להתבצע באמצעות האלגוריתם של Rocchio אשר
מעדכן את ווקטור השאילתא שהמשתמש הקליד על פי תגובת המשתמש לתוצאות
שהמנוע החזיר לשאילתא בהתאם לערכי הפרמטרים שקורטיס את רמת ההתחשבות
בשאילתת המשתמש לעומת תגובתו לתוצאות השאילתא .