

## חילוץ תתי נושאים ודירוג מחדש של תוצאות שאילתות מנועי חיפוש ע"י שימוש באתרי שיתוף סימניות ותיגוף / יוסף דיה

### תקציר

עם התפתחות רשת האינטרנט שחלה בשנים האחרונות והגידול בכמויות המידע, איתור מידע ברשת הפך לאחד השימושים העיקריים של משתמשי רשת האינטרנט. למרות השיפורים באלגוריתמי החיפוש והדירוג שבוצעו במנועי החיפוש ברשת האינטרנט עדיין נתקלים מחפשי המידע בבעיות רבות. שאילתות המידע של המשתמשים הינן, לרוב, מקיפות מספר היבטים ו/או בעלות מספר משמעויות שונות. נסיונות רבים בוצעו לטובת זיהוי המשמעויות וההיבטים עבור כל שאילתא וכן לטובת דירוג מחדש של תוצאות החיפוש בהתאם לצורכי המשתמש. עם התפתחות דור ה Web 2.0 קמו פלטפורמות רבות לטובת יצירת תוכן משתמשים ושיתופו בקרב משתמשים אחרים. אחת מהפלטפורמות האלו הינה פלטפורמת ה - Social Tagging המאפשרת למשתמשים לתייג פריטי מידע בתגיות המתארות את תוכנם ולהפיצם בקרב משתמשים אחרים. אחד מהשימושים הפופולארים בפלטפורמת ה Social Tagging הוא במערכות Social Bookmarking המאפשרות למשתמשים להעלות כתובות URL לאתר ייעודי ולתייגם בתגיות שלרוב מתארות את תוכן דפי ה URL. במחקרנו הצענו אלגוריתם ומנשק משתמש המשלבים ביו נתוני מנוע חיפוש סטנדרטי לבין נתוני מערכות Social Bookmarking לטובת חילוץ תתי נושאים עבור שאילתת חיפוש נתונה ודירוג מחדש של תוצאות החיפוש בהתאם לתתי הנושאים שנבחרו על ידי המשתמש. את תתי הנושאים המוצגים למשתמש דירגנו לפי משקל Tf-Idf המשקף את איכות תתי הנושאים. את ביצועי האלגוריתם מבחינת דירוג תוצאות החיפוש בחנו באמצעות בדיקה ממוכנת ואוטומטית שקיבלה כקלט רשימה של שאילתות חיפוש, שילבה בין נתוני מנוע החיפוש ממוכן לבין נתוני האתר Delicious, בנתה רשימת תתי נושאים עבור כל שאילתא ובחרה באופן ממוכן מספר תתי-נושאים על פיהם דורגו התוצאות מחדש. תוצאות המחקר נבחנו מול ביצועי מנוע החיפוש Google כאשר המדד, עמו בדקנו את ביצועי המנועים הינו מדד ה - MAP. מחקרנו הראה כי רשימת תתי-הנושאים המחולצים עבור השאילתות אכן מהווה רשימת תתי נושאים מבחינה ואיכותית כמו כן שיטות הדירוג שהצענו אכן מציגות שיפור בביצועי האחזור לעומת ביצועי מנוע החיפוש Google.

מספר מיון : 025.04 דיה.חל תשע"ד  
מספר מערכת : 002378896