

תקציר:

במסגרת הכתיבה המדעית התפתחה מתודולוגיה ברורה לגבי האופן בו מצטטים מחברים תכנים ממקורות חיצוניים. ואולם ציטוטים של נתונים מספריים מהווה מקרה מיוחד המציב מספר אתגרים. מחברים אינם תופשים במקרים רבים את הנתון כציטטה (quotation) ועל כן הם אינם כוללים ציון דרך מדויק למיקום הנתון בפרסום המקור כמקובל לגבי ציטטות. אספקת מראי-מקום חלקיים, במיוחד במסמכים מרובי נתונים, נובעת גם מרצון מחברים להימנע מסרבול המסמך ומהטרחה הכרוכה בכך. ואולם מצב זה מקשה על קוראי הפרסום לאתר את הנתונים בכדי לוודא את נכונותם ובכדי ללמוד על ההקשר שלהם בפרסום המקורי.

התרחבות השימוש במסדי-נתונים מקוונים באינטרנט כמקור ראשוני לנתונים מספריים, מציבה אתגרים נוספים בקשר עם ציטוט נתונים הנלקחים ממקורות אלו. הנתונים המתקבלים כתוצאה מהרצת שאילתת חיפוש שהגדיר המשתמש אינם חלק ממסמך בעל דפים ממוספרים או טבלאות ממוספרות ועל כן אין המחבר יכול לספק "מיקום מדויק" עבורם. שימוש בכתובת האינטרנט של דף אחזור הנתונים אינו מהווה פתרון הולם מכיוון שמדובר במקרים רבים בכתובת אד הוק המותאמת לפעולת האחזור הנוכחית. אתגר נוסף קשור לטיפול בעדכוני נתונים. מאגרי נתונים רבים מפרסמים רק את הנתון המעודכן האחרון. במקרה זה אין לקורא יכולת לאחזר ערך נתון שציטט המחבר ואשר עודכן לאחר מכן.

בהתייחס למקורות באינטרנט בכלל ולמסדי נתונים באינטרנט בפרט מסתפקים המדריכים לכתיבה בכללים למתן מראה מקום לגופי תוכן שלמים כגון מאמרים, דו"חות או קבצי נתונים. בכך אין מדריכי הסגנון מספקים מענה הולם לציטוט נתונים בודדים או קבוצת נתונים חלקית ממסד נתונים. כמו כן, אין הכללים המובאים במדריכי הכתיבה הללו משום התייחסות מפורשת לשאלות המתודולוגיות ולאתגרים שתוארו לעיל. התוצאה הינה מחסור בכללים ברורים ואחידים לציטוט נתונים בעיקר ממאגרי מקוונים.

רישום אחד של מראי מקום לנתונים במקבץ רחב של פרסומים יכול לאפשר ניתוח אוטומטי של תבניות ציטוט נתונים. כך למשל, בדומה לניתוח תבניות ציטוט במקרה של מאמרים, ניתן היה ללמוד על חשיבותם של מאגרי נתונים יחסית למאגרי אחרים ועל השתנות העדפות מחברים למקורות נתונים על פני הזמן. המחסור בכללי ציטוט אלה מונע ניצול פוטנציאל זה.

עבודת גמר זו בוחנת כיצד יש להתאים את ציטוט הנתונים בכדי להתמודד עם האתגרים שתוארו לעיל. במסגרת העבודה אנו מציגים מודל כולל המתאר הן את אופן הפרסום של הנתונים והן את אופן ציטוטם. האלמנט המרכזי של המודל הוא הנתון העובדתי הפשוט- הנתע"פ- המוגדר בעזרת הנושא שלו, התכונה המיוחסת לנושא, אופציונאלית, הזמן אליו הוא מתייחס, ולבסוף על ידי זמן פרסומו. המודל המוצע מתייחס לנתון כאל אלמנט בר-ציטוט עצמאי המזוהה חד-חד ערכית בעזרת הגדרה פורמאלית המשתמשת בפרטי הנתון לעיל. מרכיב זמן הפרסום של הנתון משמש להבחין בין עדכונים של אותה תופעה נמדדת המהווים כולם נתונים בני-ציטוט בפני

עצמם. בנוסף לנתון הפשוט, מגדיר המודל לציטוט נתונים גם אובייקט אבסטרקטי הנקרא מסגרת פרסומית. מסגרת פרסומית מקבצת יחדיו נתונים ומעניקה להם הקשר כולל מעבר להגדרה ולפרשנות של כל נתון בנפרד. מימוש קונקרטי של מסגרת הפרסומית יכול במקרה הפשוט להיות טבלה בודדת, ובמקרה אחר מסמך דיגיטלי רב טבלאי או מסד-נתונים שלם. בכדי להדגים את השימויות של המודל המוצע, אנו מציגים בעבודה זו מתאר לבניית מערכת לפרסום נתונים פשוטים באינטרנט. בעזרת מתאר בנייה זה יישמנו אבטיפוס למערכת פרסום וציטוט נתונים הכולל שני מאגרי נתונים המהווים סימולציה של שני מפרסמים שונים. אנו מדווחים על תוצאות השימוש באבטיפוס זה בו שימשו הנסיינים הן כקוראים של טקסטים המכילים ציטוטי נתונים הבנויים על פי עקרונות המודל, והן כמחברים המשתמשים במערכת בכדי ליצור ציטוטי נתונים מסוג זה.

מס' מיון בספריה :
025.04 מיל.מו תשס"ז
מס' מערכת בספריה :
1131148