

## מגמות בדוחות מבקר המדינה בישראל - ניתוח באמצעות Text Data Analysis

ערן רוה

בהנחיית: ד"ר רון בקרמן וד"ר הרצל כהן, אוניברסיטת חיפה

הפרסום השנתי של דוחות מבקר המדינה מעורר הד תקשורתי נרחב, המתמקד במספר פרשות בודדות, שלא תמיד מאפיינות את מלוא הנושאים שנבדקו על-ידי משרד המבקר באותה שנה (בדוח השנתי מופיעים עשרות דוחות ואלפי עמודים). השאלה הרחבה יותר הינה, מה היקף הנושאים שמבקר המדינה בודק? במה הוא מתמקד? ומה משתנה במגמות אלו לאורך השנים, ועם חילופי מבקרי המדינה? מחקרים איכותניים, על ביקורת המדינה בישראל, הצביעו על מגמות שונות: מעבר מביקורת מסורתית, שבמוקדה תקינות פעולת הממשל, לביקורת 'מתקדמת' של חיסכון, יעילות, ואפקטיביות (גייטס 1998), תמורות אקטיביסטיות בדוחות מבקרי המדינה לאורך השנים (חולתא 2012). למחקר שתי מטרות מקבילות: הראשונה – לאפיין מגמות בדוחות מבקר המדינה לאורך השנים, השנייה – פיתוח מתודולוגיה מחקרית שתאפשר אפיון המגמות, באמצעות מחקר כמותי ולא איכותני, המתבסס על יכולות בתחום ניתוח טקסט (Text Data Analysis), מתוך בסיס נתונים גדול של מסמכים בשפה העברית. המחקר התבצע במספר שלבים עוקבים: שלב 1 - הורדת קבצי הנתונים מתוך ספריית מבקר המדינה, וסידור הטקסט במבנה שיאפשר חקירתו; שלב 2 - בניית מילון של הערכים השכיחים (ה"שפה") המופיעים בדוחות מבקר המדינה (כשבסיס הנתונים של ויקיפדיה משמש כמילון של ערכים רלוונטיים לשפה העברית); שלב 3 - מציאת הנוסחה המיטבית, לביצוע קלסיפיקציה של דוחות מבקר המדינה, בהתבסס על מאפיינים של הערכים בתוך המסמכים (שכיחות במסמך ביחס לשכיחות כללית במסמכים); שלב 4 - ניתוח הנתונים באמצעות כלי ויזואליזציה של המידע (Data Visualization). תרומת המחקר היא בפיתוח מתודולוגיה כללית לניתוח כמותי ויזואלי, של כל המסמכים הנכללים בבסיס מידע גדול, בשפה העברית, שדרך כלל נחקרו באופן איכותני מדגמי.