

## האם ניתן ללמוד מתוך התוכן הקיים במדיה החברתית על הצרכים של משתמשי הדרך?

אילת גל-צור, צבי קופליק, עינת מינקוב, איתי שור, סוזן גרנט-מולר  
וסילביו נוצ'רה

### תקציר

למידע יש תפקיד מכריע בתהליכי קבלת ההחלטות בתחומי תכנון התחבורה, ניהול התנועה וקידום יעדי מערכת התחבורה. מקומה של המדיה החברתית כמקור מידע חדשני, היכול להעשיר ולהשלים מקורות מסורתיים, קיבל הכרה בתחומי חיים רבים. כפועל יוצא, מאמצי מחקר מוכוונים בשנים האחרונות לבחינת הפוטנציאל הטמון בזירת המדיה החברתית לשיפור המידע המשמש את מתכנני התחבורה, מנהלי התנועה וקובעי מדיניות התחבורה, והתוצאות הראשוניות המתקבלות ממחקרים אלה מספקות ראיות מעודדות ליכולת לנצל פוטנציאל זה. בו בזמן, עדיין קיימים אתגרים רבים הדורשים מענה על מנת להבטיח את מהימנות המידע המופק מהתוכן של מדיה חברתית, בפרט לאור האופי הבלתי מובנה של מרבית התוכן האגור במדיה החברתית.

מאמר זה סוקר את הקטגוריות השונות של מדיה חברתית, מאפייני התוכן האגור בה, וכיצד מאפיינים אלה משתקפים בתוכן הנוגע למערכת התחבורה ובאופן הגדרת המטרות המנחות את תהליך הפקת המידע החברתית. דוגמאות מאתרי מדיה חברתית בארץ משולבות במאמר על מנת לבסס את ההכרה כי תחום זה רלוונטי גם למקבלי ההחלטות בתחום התחבורה בישראל.

הספרות העדכנית בתחום נסקרת, ומקרה בוחן המתמקד בהפקת מידע מציוצים של נוסעים למשחקי כדורגל של קבוצת ליוורפול מתואר ביתר פירוט. מקרה בוחן זה מבוסס על שילוב בין הבנת הצרכים בתחום התחבורה ובין הטכניקות הקיימות בתחום כריית המלל. התוצאות מעידות על היכולת להפיק נפח משמעותי של מידע הרלוונטי לקבלת החלטות תחברתיות.

על מנת לחשוף את קצה הקרחון של הפערים שעדיין יש לסגור על מנת לנצל את מלוא הפוטנציאל של המדיה החברתית כמקור מידע לשיפור תהליכי קבלת ההחלטות בתחבורה, נסקרים האתגרים שגישה זו מציבה ובפרט בתחום האונטולוגיות, הניתוח הרגשי וזיהוי המיקום. תרומתו של מאמר זה הוא בפריסת התרומה האפשרית והאתגרים הטמונים בהפקת מידע תחברתי ממדיה חברתית ובאיתור הכוונים להמשך המחקר בתחום זה.

## הקדמה

בשנים האחרונות התבסס מקומה של המדיה החברתית כזירה באמצעותה מיליוני אנשים בארץ ובעולם משתפים את דעותיהם, מאווייהם והאירועים בחייהם עם מעגל החברים שלהם וגם עם הציבור הרחב. זירה זו הפכה למקור מידע לגבי אירועים שגרתיים וחריגים ולגבי רחשי הלב של הציבור בתחומי חיים שונים. דוגמאות לתופעה זו אפשר למצוא בשימוש במדיה חברתית כמקור לחיזוי תמיכה פוליטית כתחליף לסקרים (Tumasjan et al., 2010), ניתוח העדפות הציבור ביחס למוצרים ספציפיים (Kushal et al., 2003; Pang et al., 2008) וניטור מדדים של בריאות הציבור לצורך זיהוי התפרצות של מחלות (Grishman et al., 2002; Corley et al., 2010).

בספרות קיימות הגדרות רבות של מדיה חברתית, ולהלן שתיים מהן :

A group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 that allow the creation and exchange of User Generated Content [Kaplan et al., 2010]. An online environment where content is created, consumed, promoted, distributed, discovered or shared for purposes that are primarily related to communities and social activities, rather than functional, task-oriented objectives [Gartner Inc.<sup>1</sup> described it as].

שני מאפיינים עיקריים משותפים לשתי הגדרות אלה, וכן להגדרות רבות אחרות של מדיה חברתית. המאפיין הראשון והחשוב ביותר הוא שהתוכן נוצר ונצרך על ידי משתמשים, קרי כלל הציבור, בניגוד לתוכן הנוצר על ידי ארגונים. המאפיין השני מתייחס לכך שמדובר במגוון של אתרים ואפליקציות. שבע קטגוריות עיקריות של מדיה חברתית הוגדרו על ידי Sterne (2010) :

- Forums and Messages Boards : אתרים/אפליקציות שייעודם העיקרי הוא שאלות, תשובות ודעות, לדוגמא פורום "בונים בית"<sup>2</sup>.
  - Review and Opinion Sites : אתרים/אפליקציות שייעודם העיקרי הוא שיתוף חוות דעת והמלצות, לדוגמא tripadvisor.<sup>3</sup>
  - Social Networks : אתרים/אפליקציות שייעודם העיקרי לאפשר תקשורת בין אישית תוך שמירה על מידה מסוימת של פרטיות, לדוגמא Facebook.<sup>4</sup>
  - Blogging : אתרים/אפליקציות שייעודם העיקרי לשתף רעיונות ודעות ברמת העמקה מסוימת ובפרט כאלה המצריכים כמות גדולה יחסית של מלל, לדוגמא blogger.<sup>5</sup>
  - Micro Blogging : אתרים/אפליקציות שייעודם שיתוף מסרים קצרים, לדוגמא Twitter.<sup>6</sup>
  - Bookmarking : אתרים/אפליקציות שייעודם תיוג פריטים על פי מידת העניין שלהם למשתמש, לדוגמא reddit.<sup>7</sup>
  - Media Sharing : אתרים/אפליקציות שייעודם שיתוף תוכן שאיננו מילולי, קרי תמונות וקטעי וידאו, לדוגמא YouTube.<sup>8</sup>
- מטבע הדברים, האופקים החדשים שפתחה המדיה החברתית בפני מקבלי החלטות בתחומי חיים שונים לא פסחה גם על מקבלי החלטות בתחום

התחבורה. החשיבות של מידע בכלל, ושל מידע על צרכי והעדפות משתמשי מערכת התחבורה בפרט, אינה מוטלת בספק. התוכן האגור במדיה חברתית יכול לשמש כמקור למידע בזמן קרוב לאמיתי לגבי זמני נסיעה, ביקושים ברשת התחבורה וזיהוי אירועים חריגים, מידע הנדרש לצורך קבלת החלטות ניהול תנועה ברמה הטקטית (Mai et al., 2013; Pender et al., 2014). יתירה מכך, ניתוח מידע המופק ממדיה חברתית על פני זמן, ונוגע לשינויים בדפוסי הביקוש והעדפות משתמשי הדרך, הינו בעל פוטנציאל לתמוך בהחלטות אסטרטגיות לקידום יעדי מערכת התחבורה (Schweitzer, 2012; Collins et al., 2013 Carrasco et al., 2008).

בצד הפוטנציאל הטמון במדיה החברתית כמאגר מידע ייחודי וזמין, האתגרים להפיקו למידע ממוקד ומהימן, היכול לשמש כקלט לקבלת החלטות, רבים ומורכבים. הפרדת המוץ מן התבן, קרי איתור התוכן הרלוונטי והבנת המשמעויות שלו, מחייבים שימוש מושכל במתודולוגיות קיימות ופיתוח מתודולוגיות מתקדמות יותר בתחום כריית המלל (Text Mining). רשויות התחבורה בארץ ובעולם עושות את צעדיהן הראשונים בהפגנת נוכחות בזירת המדיה החברתית, ובניצול זירה זו להבנת צרכי משתמשי הדרך ולמתן מענה נאות לצרכים אלה.

בישראל, כמו בעולם, משתמשי מערכת התחבורה פעילים בזירת המדיה החברתית, ולהלן שתי דוגמאות מתוך רבות לתופעה זו:

ב-11 בפברואר 2015 נשלח ציוץ בנוסח "רק בבני ברק תחנת אוטובוס, מדחן ואדום לבן"<sup>9</sup> מלווה בצילום של מרחב תחנת אוטובוס שבתחומו סימון המדרכה אוסר חניה ובו בזמן מוצב בו מדחן. הציוץ זכה כמובן לציוצים אוהדים מאזרחים, אך לא לתגובות מי מרשויות התמרור.

ב-8 במאי 2014 נשלח ציוץ בנוסח "ככה צריך להיות אוטובוס בישראל - מקום להטענת הסלולרי ורדיו אישי מובנה. אוטובוס מטרופולין הבוקר. אגד- הגיע הזמן שתתקדמו"<sup>10</sup> והיווה "יריית פתיחה" לחילופי ציוצים בין מספר משתמשי Twitter.

דוגמאות מסוג זה מגלות את קצה הקרחון של התוכן המגוון הנוגע למערכת התחבורה ואגור באתרים ואפליקציות שונות השייכות למדיה החברתית. מטרתו של מאמר זה הינה לסקור את חזית הידע בתחום ההזדמנויות והאתגרים העומדים בפני אנשי המקצוע בתחום התחבורה, המבקשים לנצל את תכני המדיה החברתית כמקור מידע לשיפור תהליכי קבלת החלטות בניהול תנועה ובתכנון תחבורה.

בפרק 2 של מאמר זה מתואר התפקיד הייחודי שיש למידע ממדיה חברתית בתהליכי קבלת החלטות בתחום התחבורה. חזית הידע בשימוש בטכניקות כריית מלל בתחום התחבורה מתוארת בפרק 3. פרק 4 מוקדש לתיאור מחקר שכלל מקרה בוחן להפקת מידע תחבורתי מציוצים של אוהדי קבוצת הכדורגל של ליוורפול. כמה מן האתגרים העיקריים שעדיין מוצבים בדרך למיצוי הפוטנציאל של מדיה חברתית כמקור מידע מוצגים בפרק 5. סיכום ומסקנות נתונים בפרק 6.

### **התפקיד הייחודי של מידע ממדיה חברתית לצרכים של תכנון תחבורה וניהול תנועה**

איכות תהליכי קבלת ההחלטות בתחום התחבורה, הן ברמה האסטרטגית והן ברמה הטקטית, תלויה במידה רבה באיכות המידע המשמש קלט לתהליכים אלה. היקף המידע, שלמותו, עדכניותו ורמת הדיוק שלו הם פרמטרים משמעותיים בתהליכים אלה. בו בזמן, המידע התחבורתי משמש לקבלת החלטות של משתמשי הדרך בבואם לקבוע האם, מתי וכיצד לבצע נסיעה (Nocera, 2011). אי לכך, רשויות התחבורה משקיעות משאבים באיסוף והפצת מידע, תוך שימוש בערוצים שונים. הן מקורות איסוף המידע והן האמצעים להפצתו כוללים התקנים פיזיים ברשת הדרכים (גלאים, מצלמות, שילוט מידע אלקטרוני וכד'), הן מקורות ואמצעים דיגיטליים (ניטור GPS, שימוש באתרי אינטרנט וכד') והן אמצעים ידניים ובפרט שאלונים מסוגים שונים. למרות ההתפתחויות בתחום איסוף ועיבוד המידע הדיגיטלי המבוסס על ניטור התקנים הנישאים על ידי משתמשי הדרך (ניטור מבוסס סולר, ניטור באמצעות גלאי Bluetooth וכד'), התובנות שניתן להפיק ממנו מוגבלות, בפרט בנוגע למוטיבציה ולתפיסה הסובייקטיבית של משתמשי הדרך בפעילות הניידות שלהם. מסיבה זו, תפקידם של השאלונים לא נפקד מהמערך הכולל של איסוף מידע המשמש לקבלת החלטות בתחום התחבורה.

עיבוד מידע מובנה, קרי מידע שפרטי הנתונים המרכיבים אותו מוגדרים מראש (כגון מידע ממערכות GPS), הוא עיבוד פשוט יחסית. עם זאת, מומחים מערכים כי כ- 80% מהמידע המופץ ברשת האינטרנט בכלל ובמדיה חברתית בפרט, הוא מידע בלתי מובנה (Lui et al., 2011), ולכן התמקדות במידע המובנה בלבד משמעותה התעלמות מהרוב המכריע של התוכן הנמצא בזירת תקשורת זו, תוכן שיכול להיות ערך מוסף מהסוג המושג כיום על ידי שאלונים. אי לכך, אין זה מפתיע כי מאמצי מחקר לא מעטים מופנים לפיתוח מתודולוגיות וטכניקות לניתוח של מידע בלתי מובנה, ובעיקר למידע הכולל מלל חופשי.

האתגר העומד בפני מקבלי ההחלטות בתחבורה הוא האם וכיצד ניתן לעשות שימוש במדיה חברתית, כמשלים (או מחליף) את מקורות המידע הקיימים, וכמייצר הזדמנות לשיפור איכות המידע ולהגדלת היעילות של הפצתו, תוך עמידה במבחן עלות-תועלת.

על מנת לענות על שאלה זו, יש לבחון תחילה את מאפייני התוכן הקיים במדיה חברתית בכלל, ואת התוכן הנוגע לתחבורה בפרט. קיימים שלושה מאפיינים עיקריים לתוכן האגור במדיה חברתית ונוצר על ידי משתמש פרטי (בניגוד לארגון) (Gal-Tzur et al., 2014a):

- התוכן מתייחס בדרך כלל לאירוע שהמשתמש חווה או לפעולה שבכוונתו לבצע.
- האירוע או הפעולה שהם בליבת התוכן התרחשו מעט לפני או יתרחשו מעט אחרי המועד בו המשתמש יצר את התוכן.
- האירוע או הפעולה שהם בליבת התוכן הינם בעלי חשיבות מסוימת למשתמש. מאפיין זה הינו בעל חשיבות מיוחדת, מאחר ובניגוד לשאלון, הנותן ביטוי בעיקר לנושאים הנתפסים כמשמעותיים מנקודת

המבט של אנשי המקצוע בתחבורה, תוכן המופץ ביוזמת משתמשי הדרך מעיד על הנושאים הנתפסים כחשובים מנקודת מבטם.

מאפיינים אלה מהווים את הבסיס להגדרת המטרות של תהליך הפקת המידע הרלוונטי לתכנון ולשיפור יכולת הניידות והנגישות של משתמשי מערכת התחבורה. המידע נוגע לאמצעי תחבורה, כגון רכב פרטי, תחבורה ציבורית או אופניים, וכן לתשתיות התחבורה כגון כבישים, מגרשי חניה או תחנות השכרת אופניים. בנוסף, המידע נוגע לשירותים הנדרשים לצורך פעילות הניידות והנגישות כגון תכנון נסיעה, תשלום על אמצעי נסיעה או הזמנת שירותי נסיעה.

מתוך תובנות אלה ניתן להגדיר שלושה סוגי מידע המהווים את יעדי תהליך הפקת המידע התחבורתי מתוך המדיה חברתית:

1. מידע על הצורך בביצוע נסיעה ממוצא ליעד, בדומה למסר שפרסם משתמש Facebook:

"אני צריך להגיע מירושלים לתל-אביב במוצאי שבת. אני צריך להיות בדרום תל-אביב בשבע וחצי. לא יהיה לי רכב באותו היום. התחבורה הציבורית מתחילה לפעול אחרי שמונה. איך אני יכול לעקוף את התאוקרטיה הזאת? יש, למשל, מוניות שירות כלשהן? לעזרתכם אודה!"<sup>11</sup>

2. זיהוי אירועים חריגים הקשורים לרשת התחבורה או לשירותי תחבורה, בדומה לציוץ:

"על איילון דרום מגיעה מסופשבוע בצפון, הכביש חסום, מההפגנה... לא רגילה להיות בצד של המכוניות. זה עוד יקח הרבה זמן? נשמח פה למידע פנימי!"<sup>12</sup>

3. הבעת דעה של משתמשי מערכת התחבורה בנוגע לאיכות שירותי תחבורה, בדומה למסר שפרסמה נוסעת באתר "תחבורה בדרך שלנו", בלוג המידע של משתמשי התחבורה הציבורית בישראל:

"...קודם כל אציין שיש כמה וכמה תחנות אוטובוס מזעזעות, אקח לדוגמה את תחנת האוטובוס ליד קניון אבנת. כמה מביניכם שגרים בפ"ת בטח יודעים למה אני מתכוונת. זוהי תחנה בלוייה, ללא כיסאות, שבורה, ממש לא עושה חשק לעמוד (אין איפה לשבת!) ולחכות לאוטובוס שם..."<sup>13</sup>

שלושת סוגי המידע הנ"ל מספקים את הבסיס הראשוני לפיתוח ויישום מתודולוגיות להפקת מידע תחבורתי מתוך עושר התוכן הקיים במדיה חברתית.

### חזית הידע בתחום השימוש בכריית המלל לצרכי תחבורה

בשנים האחרונות חוקרים בתחום התחבורה החלו לסלול את הדרך להתמודדות עם האתגרים של שימוש בשיטות של כריית מלל לניתוח תוכן בעל הקשר תחבורתי מתוך מדיה חברתית. כריית מלל הינו תהליך העושה שימוש במגוון טכניקות וכלים, שיש להתאימן היטב למשימת החיפוש עבורה הם מיושמים.

סקירת חזית הידע הנוגעת ליישום של שיטות כריית מלל לשימושי תחבורה מגלה מגוון תחומים בהם חוקרים שונים בחרו להתמקד.

Gao et al (2013) הדגימו את הפוטנציאל של טכניקות כריית מלל לניתוח דוחי תאונות דרכים במטרה לסווג סוגי תאונות. Mai and Hranac (2013) חקרו את התאימות בין ציוצי משתמשים בנוגע לאירועי תנועה חריגים לבין תיעוד אירועים מסוג זה שבוצעו על ידי רשות התחבורה. החוקרים התמקדו בציוצים שכללו תיעוד גיאוגרפי והשתמשו באנליזה סמנטית, קרי זיהוי מונחים רלוונטיים במלל של הציוצים, כאשר לכל מונח הוצמד ערך המשקף את מידת הרלוונטיות שלו לאירוע תחבורתי, וכל ציוץ קיבל דירוג כולל המאפשר לזהות את אלה שהינם בעלי סבירות גבוהה לקשר לאירועי תחבורה חריגים. נמצא מתאם טוב בין הציוצים שזוהו לבין דוחי רשות התחבורה על אירועים מסוג זה. גם Schulz et al. (2013) התמקדו במידע מ-Twitter להפקת מידע הנוגע לסביבה בה התרחשו תאונות. החוקרים הראו כי המידע שהופק הניב ערך מוסף לגבי הבנת מאפייני המצב יחסית למידע הפורמלי שניתן היה לקבל ממקורות אחרים, וכי הוא מאפשר ניהול טוב יותר של האירוע החריג.

Schweitzer (2012) יישמו טכניקות של כריית מידע להבנת תפיסתם של משתמשים את רמת השירות של שירותי תחבורה שונים, כולל תחבורה אווירית. גם מחקר זה עשה שימוש ב-Twitter כמקור מידע, והביא ראיות לכך שניתן לזהות מגמות של שביעות רצון באמצעות הדגמת המתאם בין המידע שהופק לבין אירועים חריגים, כגון גל חום שהשפיע על נוחות הנסיעה בתחבורה ציבורית. מטרה דומה ואותו מקור מידע שימשו גם את Collins et al (2013) במחקרם שעסק ברמת שביעות רצון של משתמשי תחבורה ציבורית. גם הם הראו ממצאים שהופקו מהציוצים עלו בקנה אחד עם אירועים חריגים, כדוגמת נפילת חשמל בתחנת רכבת שגרמה לאיחורים חריגים בתנועת הרכבות. החוקרים מצאו כי משתמשי התחבורה הציבורית נוטים לציץ יותר לגבי אי שביעות רצון מרמת השירות בהשוואה להבעת דעות חיוביות.

הפקת מידע על מסלולי נסיעה בתחבורה ציבורית היא המטרה שעמדה בבסיס מחקרם של Steiger et al. (2014). החוקרים עשו שימוש במידע משולב שכלל תוכן מילולי מ-Twitter וכן תמונות מתוייגות מ-Flickr ומ-Instagram. שיטות של ניתוח מרחבי וניתוח מילולי יושמו על מקרה בוחר בלונדון, והמחקר הדגים כי מתוך מקורות מידע אלה ניתן לזהות מסופי תחבורה ציבורית ומסלולי נסיעה בתחבורה ציבורית.

מחקרים אלה מספקים בסיס טוב להנחה כי ניתן להפיק מידע בעל ערך תחבורתי מתוך המדיה החברתית, וכי יש מקום להמשיך ולפתח מתודולוגיות וטכניקות למטרה זו. בפרק הבא מובא תיאור מפורט יותר של מחקר שמטרתו לבחון את היכולת לעשות שימוש בטכניקות של כריית מלל להפקת מידע תחבורתי מגוון, הנוגע לצרכים והעדפות של משתמשי מערכת התחבורה.

### **הדגמת הפוטנציאל בתחום התחבורה**

בפרק 2 של מאמר זה הוגדרו שלושה סוגי מידע המהווים את יעדי תהליך הפקת המידע התחבורתי מתוך המדיה החברתית: (1) מידע על הצורך בביצוע

נסיעה ממוצא ליעד, (2) זיהוי אירועים חריגים הקשורים לרשת התחבורה או לשירותי תחבורה, (3) הבעת דעה של משתמשי מערכת התחבורה בנוגע לאיכות שירותי תחבורה. על מנת להדגים את ההתכנות של השגת יעדים אלה, הוגדר מקרה בוחן המתמקד בניסיון להבין את הצרכים וההעדפות בהקשר של שירותי תחבורה של אוהדי קבוצת הכדורגל ליוורפול, והוא מתואר בפירוט במאמרם של Gal Tzur et al (2014a). מן הראוי להדגיש כי מטרתו של מקרה הבוחן הייתה בעיקר לבחון את ההתכנות של הפקת מגוון סוגי מידע בעלי חשיבות לתהליכי קבלת ההחלטות בתחום התחבורה, תוך שימוש באותו מקור מידע ובאותה טכניקה להפקתו. התובנות לגבי מאפייני הניידות והעדפות הנסיעה של אוהדי קבוצת ליוורפול מספקות ערך מוסף מעבר למטרה עיקרית זו.

הבחירה בקבוצת משתמשים זו לא היה מקרי. ראשית, מדובר בקבוצה פעילה מאד ברשת ה-Twitter, רשת חברתית המספקת ממשק פתוח להורדת המסרים הנשלחים בה. מחקר שבוצע על ידי Lanagan et al. (2011) הראה כי יכול לשמש כמקור מהמון למידע הקשור לאירועי כדורגל. הסיבה השנייה לבחירה נובעת מכמה מאפיינים ספציפיים של אוהדי קבוצת הכדורגל, מאפיינים המקלים על זיהוי המסרים של קבוצה זו. בפרט, מועדי ומיקום המשחקים ידוע, ולכן מועד ביצוע הנסיעות ויעדן (בזמן ההגעה למשחק) או מוצאן (בעת סיומו) ידוע. הסיבה השלישית למיקוד בציודי אוהדי ליוורפול הייתה השפה הבינלאומית (אנגלית) בה הם כתובים.

#### - שלבי תהליך כריית המלל

השלב הראשון בתהליך כריית המלל התמקד בזיהוי ציוצים אשר קשורים למשחקי הכדורגל של ליוורפול. מקרה הבוחן התמקד בשלושה משחקים, והחיפוש התמקד בציוצים שנשלחו בשבוע בו התקיים המשחק, החל מ-3 ימים לפני המועד בו התקיים. המונחים המילוליים שהנחו את תהליך סינון היו אלה שרלוונטיים למשחקי הקבוצה כגון שם מועדון האוהדים של הקבוצה, שמות האצטדיונים בהם התקיימו המשחקים וכד'. בשלב זה המטרה הייתה לאתר את כל הציוצים הרלוונטיים למשחקים, גם ב"מחיר" של זיהוי שגוי של ציוצים שאינם רלוונטיים למטרת תהליך החיפוש, מאחר וההנחה הייתה כי ציוצים שגויים ינופו בשלבים הבאים. בסיומו של שלב זה נבנה קורפוס (בסיס נתוני מלל) של 3 מיליון ציוצים.

ברור כי רק חלק קטן מתוך מיליוני הציוצים שנשלפו אכן רלוונטיים לתחבורה. בשלב השני, במטרה לאתר ציוצים שקימת סבירות גבוהה להיותם רלוונטיים לתחבורה, נבנה לקסיקון של מונחי תחבורה. בהיעדר אונטולוגיה מתאימה בתחום זה (פירוט יתר בנושא אונטולוגיות בתחבורה נתון בפרק הבא של מאמר זה), נבנה הלקסיקון בשיטה אוטומטית למחצה. 35 מסמכים בתחום התחבורה, הכוללים מאמרים מדעיים, אתרי אינטרנט ופורומים של משתמשים, נסקרו אוטומטית על מנת לזהות מילים וביטויים חוזרים ונשנים במסמכים אלה, ואינם נמצאים במסמכים מתחומים אחרים. למילים וביטויים חוזרים הוצמדו ערכים המשקפים את כמות הפעמים שכל אחד מהם הופיע במסמכים שנבדקו, ורשימת המילים והביטויים השכיחים הועברה למומחי תחבורה אשר סיפקו הערכה מקצועית-סובייקטיבית למידת הרלוונטיות שלהם לתחבורה. הלקסיקון הסופי כלל כ-500 מונחים שקיבלו

ניקוד של 3-5 (בסקאלה של 0-5) בהערכת מומחי התחבורה, והכיל מונחים כגון "passengers", "vehicle", "congestion", "bus", "traffic". לקסיקון זה שימש לדירוג כל אחד משלושת מיליון הציוצים שנשלפו בשלב הראשון, כאשר הניקוד שקיבל כל ציוץ חושב כסכום הניקוד של מונחי הלקסיקון שהופיעו בו.

1500 הציוצים שקיבלו את הניקוד הגבוה ביותר שימשו כקלט לשלב כריית המלל השלישי. בשלב זה נעשה שימוש בשיטות של למידה מונחית לבניית שני סוגים של כללי סווג, כללים המפרידים בין ציוצים שנשלחו על ידי רשויות לבין כאלה שנשלחו על ידי אנשים פרטיים וכללים המפרידים בין ציוצים הקשורים לתחבורה וכאלה שלא.

לרשויות תחבורה רבות בעולם בכלל ובבריטניה בפרט יש חשבון Twitter באמצעותו הן שולחות למשתמשים הודעות לגבי פעולות ואירועים במערכת התחבורה. מאחר והמטרה של מקרה הבוחן היא ללמוד על צרכי והעדפות משתמשי מערכת התחבורה, יש לסנן ציוצים שמקורם ברשויות וארגונים אחרים (כגון מפעילי תחבורה ציבורית). למידה מונחית, בהקשר של בניית כללי סווג למקור הציוץ (ארגון או משתמש פרטי), הינה תהליך בו מתבצע תחילה סווג ידני של ציוצים כנוצרים על ידי ארגון או על ידי משתמש פרטי. לאחר מכן מתבצעת בנייה אוטומטית של כללי סווג, המתבססת על דוגמאות מסווגות אלה ועל מציאת המאפיינים המפרידים בין דוגמאות שסווגו ככאלה שנשלחו על ידי רשות/ארגון וככאלה שנשלחו על ידי משתמשי מערכת התחבורה. מספר שיטות למידה מונחית נבחנו, והשיטה שנמצאה כיעילה ביותר במקרה זה, ושימשה לקבלת התוצאות הסופיות הינה שיטת סווג (Steinwart et al., 2008) SVM.

כללי הסיווג שהופקו בתהליך כריית המלל למקור הציוץ כללו מגוון תנאים. לדוגמה, נמצא כי שמות כגון "I" או "we", וכן מונחים בלתי פורמליים כגון "lol" מעידים על כך שהציוץ נשלח על ידי משתמש פרטי. מאידך, המילה "due" נמצאה כמסווגת היטב ציוץ שמקורו ברשות או ארגון, כנראה מאחר וארגונים נוהגים להשתמש בה לצורך הסבר על הסיבות לאירועים ברשת התחבורה.

בהתבסס על כללי הסיווג של ציוצי משתמשים פרטיים, נבנה בסיס דוגמאות נוסף של 1500 דוגמאות. התיוג הידני של דוגמאות אלה כשייכות או לא שייכות לתחבורה בוצע על ידי שלושה מומחי תחבורה. בהשוואה בין זוגות תיוגים של מומחים שונים נמצא ממוצע התאמה של 81%. ממצא זה מעיד על מידה מסוימת של סובייקטיביות בזיהוי, תופעה המקשה כמובן על תהליך הזיהוי האוטומטי. בניית כללי סווג אוטומטיים בתהליך של למידה מונחית הפיק כללים המשקפים שילוב של מונחים תחבורתיים (כגון "bus" ו-"station") עם מילות יחס רלוונטיות כגון "from".

הוכחת ההתכנות של תהליך כריית מלל חיבת להיבחן על בסיס מדדים כמותיים. קיימים שני מדדי הצלחה מקובלים לבחינת איכות הסיווג של תהליך כריית המלל. דיוק (Precision) מודד את רמת הדיוק של המסרים שסווגו באמצעות תהליך כריית המלל, על בסיס היחס בין מספר המסרים שסווגו נכונה לבין מספר המסרים הכולל שסווגו כשייכים לקטגוריה. אחזור (Recall) מודד את רמת היכולת של תהליך כריית המלל לאתר את כלל המסרים השייכים בפועל לקטגוריה הנבחנת, ומחושב כיחס בין מספר



המסרים שסווגו נכונה לבין מספר המסרים הכולל שאכן שייכים לקטגוריה זו. המשלים של מדד הדיוק מקביל לטעות מסוג ראשון בסטטיסטיקה, והמשלים של מדד האחזור מקביל לטעות מסוג שני. קימת כמובן תלות בין המדדים, ושיפור באחד מהם מוביל בדרך כלל להחמרה באחר. בטבלה 1 נתונים ערכי מדדי ההצלחה של שני סוגי כללי הסיווג שפותחו. ערכים אלה נחשבים טובים בסטנדרטים של תחום כריית המלל.

**טבלה 1: מדדי ההצלחה של כללי הסיווג למקור הציוץ ולרלוונטיות של לתחבורה**

מדד האחזור (Recall)	מדד הדיוק (Precision)	מטרת כללי הסיווג
0.92	0.88	זיהוי מסרים שנשלחו על ידי משתמשים פרטיים
0.78	0.8	זיהוי מסרים הקשורים לתחבורה

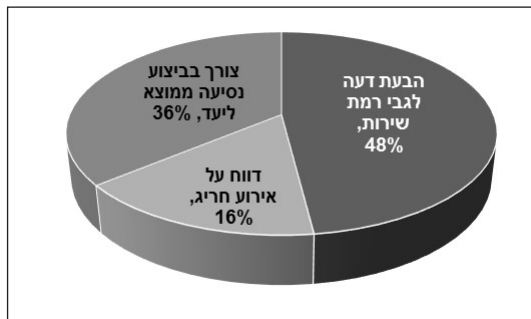
בשלב הרביעי של תהליך כריית המלל, פותחו כללי סיווג להפרדה בין שלושה סוגים של מסרים הנוגעים לתחבורה, בהתאם לשלושת סוגי המידע התחבורתי שהוגדרו כיעדי החיפוש (הצורך בביצוע נסיעה ממוצא ליעד, זיהוי אירועים חריגים והבעת דעה כל רמת השירות). גם במקרה זה נבנה בסיס דוגמאות ייעודי למטרה זו, ושני מומחי תחבורה ביצעו תיוג ידני של כל ציוץ כמבטא אחד משלושת סוגי המידע האפשריים. גם במקרה זה ההסכמה בין המומחים לא הייתה מלאה ונעה בין 78% לציוצים הנוגעים לצורך בביצוע נסיעה ל-94% לגבי ציוצים העוסקים בדווח על אירוע חריג. הסיווג לשלוש קטגוריות של סוגי מידע הינו עדין יותר מההפרדה בין ציוצים רלוונטיים לתחבורה לכאלה שלא. מדדי ההצלחה שהושגו אכן משקפים קושי זה, והם נתונים בטבלה 2. ניתן לראות כי ציוצים הכוללים הבעת דעה על רמת השירות סווגו באופן מהימן יותר, כנראה בזכות מילים מובהקות כגון "excellent" או "terrible" המאפיינות מסרים מסוג זה. מאידך, איכות הזיהוי של ציוצים הנוגעים לצורך בביצוע נסיעה הייתה הנמוכה ביותר, כנראה בשל האופנים הרבים לביטוי צורך זה, שלעיתים אינו מבוטא במפורש אלא במשתמע. ואכן, גם ההסכמה בין התיוג האנושי של מידע מסוג זה הייתה הנמוכה ביותר. מידת ההצלחה בסיווג מסרים הנוגעים לאירוע חריג הייתה גם היא נמוכה יחסית. האתגרים בכריית מלל בכלל, ומלל תחבורתי בפרט, המתוארים בפרק 5 להלן מהווים הסבר מסויים לצורך בהעמקת המחקר ופיתוח מתודולוגיות נוספות לביצוע מטלה מורכבת זו.

**טבלה 2: מדדי ההצלחה של כללי הסיווג לסוג המידע התחבורתי**

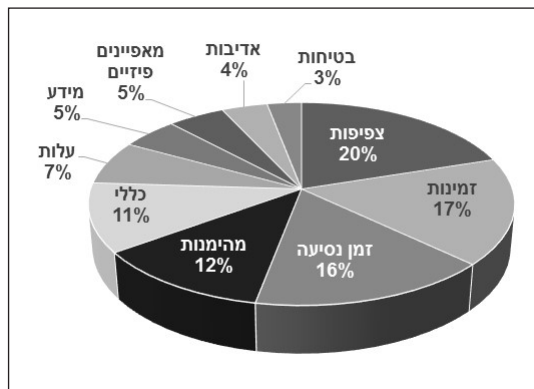
מדד האחזור (Recall)	מדד הדיוק (Precision)	סוג המידע
0.58	0.53	הצורך בביצוע נסיעה ממוצא ליעד
0.6	0.63	זיהוי אירועים חריגים
0.79	0.68	הבעת דעה כל רמת השירות

**- ממצאים מניתוח ציוצי אוהדי קבוצת הכדורגל של ליוורפול**

מעבר לבחינת היכולת לזיהוי אוטומטי של ציוצים העוסקים בתחבורה בכלל, וסוגי המידע האגורים בהם, הניב מקרה הבוחן כמה ממצאים מעניינים על התוכן האגור במדיה החברתית לגבי צרכי והעדפות המשתמשים. באיור 1 נתון אחוז הציוצים, מתוך אלה שנמצאו רלוונטיים לתחבורה, המתמקדים בכל אחד מסוגי המידע שנכללו בתהליך החיפוש. ניתן לראות בבירור כי לכל שלושת הסוגים יש ייצוג נאות במדיה החברתית שנבנתה, עובדה המהווה תמריץ למקבלי ההחלטות להשתמש במקור מידע זה כקלט למגוון תהליכי תכנון תחבורה וניהול תנועה.



**איור 1:** אחוז הציוצים המשקפים כל אחד מסוגי המידע התחבורתי (Gal-Tzur et al., 2014b)



**איור 2:** אחוז הציוצים המתייחסים להיבטים השונים של רמת השירות (Gal-Tzur et al., 2014b)

הציוצים שכללו הבעת דעה לגבי היבטים שונים של רמת השירות של שירותי התחבורה, אשר הרוב המכריע שלהם התייחס לתחבורה ציבורית, הוא את הנתח הגדול ביותר מבין כלל הציוצים, ולגביהם בוצע סיווג ידני לרמת פירוט נוספת. רמת פירוט זו התייחסה להיבטים השונים של רמת השירות, והתוצאות מובאות באיור 2. ניתן לראות בבירור כי היבט הצפיפות הוא המשמעותי ביותר בתפיסת רמת השירות של אוהדי קבוצת ליוורפול. מאידך, עלות הנסיעה, היבט שניתן היה לחשוב שמהווה שיקול מהותי, זוכה רק ל-7% מכלל ההתייחסויות של אוהדי הקבוצה לרמת השירות של התחבורה הציבורית. ממצאים אלה לא רק מהווים מוטיבציה להפעלת השיטה בהקשר של משחקים נוספים של הקבוצה על מנת להרחיב את היקף הניתוח, אלא גם מספקים בסיס ראשוני לבחינת העלאת תדירות שירותי התחבורה למשחקי הכדורגל כבסיס לשיפור רמת שביעות הרצון של הנוסעים.

### האתגרים בכריית מלל בכלל ובתחום התחבורה בפרט

כריית מלל הינה משימה אתגרית כאשר היא מיושמת לכל אחד מתחומי החיים, ועם זאת קיימים כמה אתגרים המאפיינים במיוחד את תחום התחבורה.

#### - אונתולוגיה תחבורתית

אחד החסמים המרכזיים בעיבוד אוטומטי של מלל בכלל, ושל מלל בלתי פורמלי המצוי במדיה החברתית בפרט, הינו היעדר הקשר מלווה. לצורך הדגמה, על מנת להסיק כי מסר כגון "המטרונית הגיעה בזמן" הוא מסר הרלוונטי לתחבורה, יש לדעת כי למערכת הסעת ההמונים של חיפה קוראים מטרונית. אחת הדרכים היעילות לייצג ידע הנוגע להקשר של תחום חיים מסוים הוא שימוש באונתולוגיה. אונתולוגיות מהוות מסגרות מתודולוגיות להגדרה והמשגה של הקשרי מידע באמצעות מושגים ומונחים המקושרים ביניהם באמצעות יחסים מוגדרים (שייך ל... נכלל ב... וכדו'). תהליך כריית מלל, הכולל שלבים של תיוג מונחים וקישור חלקי מלל זה לזה באמצעות הגדרת היחסים ביניהם, יכול להתבצע באופן יעיל יותר כאשר נעשה שימוש באונתולוגיה הממוקדת בתחום החיים בו המלל עוסק. לדוגמא, לו היתה קימת אונתולוגיה שבה "מטרונית" הייתה ישות, וישות זו הייתה מקושרת ביחס "היא" למושג "תחבורה ציבורית", ניתן היה ביתר קלות להבין כי המסר "המטרונית הגיעה בזמן" הוא מסר הרלוונטי לתחום התחבורה.

סקירת ספרות של אונתולוגיות בתחבורה חושפת כי פותחו מספר אונתולוגיות באנגלית לתחום התחבורה, ועם זאת אף אחת מהן איננה מקיפה מספיק על מנת לספק בסיס מושלם לניתוח מלל תחבורתי שמקורו במדיה חברתית. הסיבה לכך היא שחלק מן האונתולוגיות הקיימות באנגלית התמקדו במטלות ספציפיות הקשורות לניידות כגון תכנון נסיעה (Niaraki et al., 2009), וחלקן באמצעי נסיעה ספציפיים, כגון תחבורה ציבורית או שיתוף נסיעות (Cho et al., 2013; Wang et al., 2005), ואי לכך חסרות את היכולת לשמש לניתוח מלל הנוגע למגוון פעילויות ניידות ולמגוון אמצעי נסיעה ושירותי תחבורה. יתירה מכך, האונתולוגיות הקיימות באנגלית משקפות מונחים מקצועיים פורמליים בתחום התחבורה, והתוכן הנמצא במדיה החברתית מתאפיין בשפה לא רשמית ובכינויים תלויי תרבות למושגים שונים. לדוגמא, הרכבת

התחתית בלונדון מכונה "tube", בעוד ההתייחסות לאותו אמצעי נסיעה בבוסטון מתבצעת באמצעות המונח "T".

מן האמור לעיל עולה כי קיימים פערי ידע בתחום האונתולוגיה התחבורתית באנגלית, קל וחומר בעברית. למיטב ידיעתם של כותבי מאמר זה, לא קימת אונתולוגיה תחבורתית בשפה העברית. האתגרים בזיהוי אוטומטי של מסרים הרלוונטיים לתהליכי קבלת החלטות בתחום התחבורה הקיימים בכל שפה, קיימים כמובן גם בעברית. הדוגמא הבאה היא רק אחת מיני רבות המעידה כי ביטויים מסוימים הנכללות במסר (במקרה זה "רכבת מרכז") אינם ערובה לכך כי המסר כולל מידע חיוני למקבלי החלטות בתחבורה:

"ליד רכבת מרכז יש עשרות אוהלים של מחוסרי דיור. קצת חורקת להם הדלת, והגג טיפה מטפטף. מה היית ממליץ לעשות כדי לרענן השטיח?"<sup>14</sup>

ניתוח מלל בעברית מורכב אף יותר משפות רבות אחרות. בעברית לא מנוקדת ריבוי משמעותיות לאותה מילה הוא שכיח בהרבה מתופעה זו בשפות כגון אנגלית. אי לכך, פיתוח אונתולוגיה תחבורתית בעברית, ובפרט כזו הכוללת ביטויים בלתי פורמליים וסלנג, הטיפוסיים למסרים הנכתבים על ידי משתמשים, תתרום ללא ספק לקידום היכולת למצוא מידע תחבורתי רלוונטי מתוך מדיה חברתית.

#### **- ניתוח רגשי (Sentiment Analysis)**

ניתוח רגשי (Sentiment Analysis) או כריית עמדות (Opinion Mining) הינו תהליך שמטרתו לזהות את עמדתו (חיובית או שלילית) של כותב המלל בהקשר לנושא מסוים. בד בבד עם המרכזיות של המדיה החברתית כזירת תקשורת בין אנשים, זוכה תחום זה לתשומת לב רבה (Pak et Al, 2010). חשיבותו של תהליך זה בתחום התחבורה מתבטא בעיקר במקרים בהם נדרש מידע לגבי יחס משתמשי הדרך כלפי אמצעי נסיעה או שירות תחבורה מסוים ולגבי תפיסתם את רמת השירות של אמצעי או שירות זה. ניתוח רגשי מתחיל בדרך כלל ביצירת לקסיקון של המילים והביטויים שנעשה בהם שימוש שכיח לצורך הפגנת רגש או דעה, וכל מילה או ביטוי מתווייגים כחיוביים (משמשים להפגנת רגש או דעה חיוביים) או שליליים. לדוגמא, מילים כגון "מצוין" ו"ידידותי" יתווייגו כחיוביות בעוד מילים כגון "גרוע" ו"מסובך" יתווייגו כשליליות. כמובן שתהליך הניתוח מתייחס גם לשילוב בין מונח חיובי ומונח המתאר שלילה (כגון "לא יפה") על מנת לזהות יחס שלילי. מילים, שמשמעותן המקורית שלילית, משמשות לעיתים לביטוי חיובי כגון "נורא יפה". ביטויים צינניים מהווים אתגר מיוחד בניתוח עמדות. "האוטובוס הזה ממש נפלא" יכול בקלות לבטא הן אהדה והן סלידה. הפרדה בין שתי חלופות אלה יכולה להתבצע על בסיס ניתוח המלל הקודם או העוקב למשפט זה.

הגדרת ביטויים המאפיינים יחס חיובי או שלילי בתחום מסוים מסייע למהימנות תהליך זיהוי עמדתו של הכותב (Pang et al., 2008). לדוגמא, מונח כגון "פקוק" לא יופיע במרבית הלקסיקונים העוסקים בכריית דעות, אך בתחום התחבורה על מונח זה להיכלל בלקסיקון.

הסקת מסקנות מצרפיות לגבי עמדות משתמשים לגבי אמצעי נסיעה מסוים או שירות נסיעה מסוים, המתבצעת על בסיס ניתוח מספר רב של מסרים בודדים, צריכה לקחת בחשבון גם את מקור המידע. לדוגמא, אתרים

מסוימים השייכים לזירת המדיה החברתית מיועדים ביסודם לתלוות של משתמשים (לדוגמא, Hellopeter.com), ואי לכך מוטים מראש לכוון מסוים. למרות האתגרים בניתוח רגשי, כבר הושגו הצלחות רבות ביישום תהליך זה בתחומי חיים שונים תוך שיעורי הצלחה של 80% בזיהוי (Pang et al., 2008), ואף בתחום התחבורה ישנן סנוניות חיוביות בכוון זה (Schweitzer, 2012). המשך ההשקעה בפיתוח שיטות לניתוח רגשי בתחום התחבורה בכלל, ובתת תחומים הנוגעים לשירותי תחבורה שונים בפרט, יניב ככל הנראה שיעורי הצלחה גבוהים.

#### - נתוני מיקום

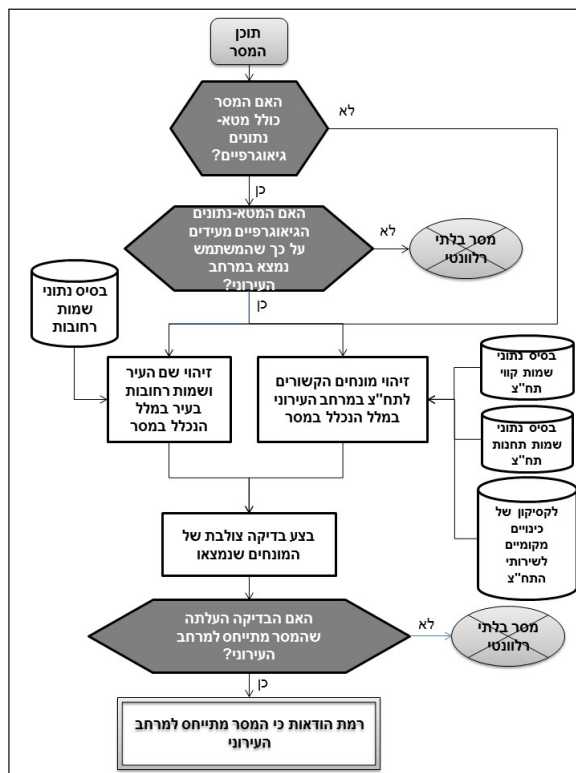
רשויות תחבורה מעוניינות במרבית המקרים לזהות מסרים רלוונטיים לתת הרשת של מערכת התחבורה הנמצאת תחת תחום אחריותם. ניתן להניח כי מרבית המסרים המופיעים באתרי אינטרנט ובחשבונות ברשתות החברתיות השייכים לרשות תחבורה מסוימת יתייחסו לנושאי ניידות במרחב הרשות. עם זאת, בשל האופי הרשתי של מערכת התחבורה (כגון כבישים וקווי תחבורה ציבורית המקשרים בין מרחבים של רשויות שונות), הפעילות התחבורתית במקור הנסיעה, במהלכה וביעדה עשויים להיות רלוונטיים לכלל רשויות התחבורה שהנסיעה מתבצעת במרחביהן. אי לכך, לצורך קבלת החלטות ברשות תחבורה מסוימת נדרש במקרים רבים ניתוח תוכן מתוך אתרי אינטרנט וחשבונות ברשתות החברתיות הן של רשויות אחרות והן של ארגונים אחרים כגון מפעילי תחבורה ציבורית, מפעילי שירותי חניה וכד', אשר מרחב פעולתם חופף או משיק עם הרשות המפיקה את המידע.

המקרה בו המסר נוצר על ידי המשתמש באתר האינטרנט או בחשבון ברשת חברתית השייך לרשות או לארגון בתחום התחבורה מקל במידה רבה על איתור תחומי המרחב ברשת התחבורה אליו מתייחס המסר. אולם, הרוב המכריע של המסרים נשלחים באמצעות אתרים ואפליקציות שאינן ספציפיות לתחבורה, ובפרט באפליקציות חברתיות כגון Facebook, Twitter ובארץ במקרים רבים באמצעות WhatsApp. במקרים אלה קיימות שתי טכניקות עיקריות לזיהוי המיקום המרחבי אליו מתייחס המסר: (1) לזהות את המיקום של המשתמש בעת שליחת המסר ו/או (2) לזהות את המיקום מתוך תוכן המסר. השיטה הראשונה מתאפשרת כאשר המסר כולל מטא-נתונים (נתונים אודות נתונים) המתייחסים למיקום הגיאוגרפי של המשתמש בעת שליחת המסר. מצב זה מתקיים כאשר האפליקציה תומכת בתכונה זו וכאשר הגדרות הפרטיות של המשתמש מאפשרות זאת.

נכון לימים אלה, משתמשים מעטים בוחרים לאפשר הצמדה של מטא-נתונים לגבי מיקומם למסרים שהם שולחים (Mai et al., 2013). מן הראוי גם לציין כי גם במידה ולמסר מוצמד מידע גיאוגרפי על המיקום בעת שליחתו, קיימים מקרים בהם המשתמש נמצא במיקום מסוים אך שולח מסר לגבי מיקום אחר. לדוגמא, משתמש שנוסע באוטובוס מירושלים לתל-אביב יכול לנהל "שיחה" ב-WhatsApp עם חבר בהיותו בתל-אביב ולספר על תחזוקה לקויה של התחנה בה עלה על האוטובוס בירושלים.

לאור המגבלות של זיהוי המיקום באמצעות מטא-נתונים גיאוגרפיים, מתבצעים מחקרים המתמקדים בפיתוח מתודולוגיות להסקת מסקנות לגבי המרחב הגיאוגרפי אליו מתייחסים מסרים במדיה חברתית על בסיס ניתוח

המלל הנכלל במסר. משימה זו אתגרית במיוחד לאור מגוון המשמעויות שניתן לייחס לשמות מקומות. לדוגמא, הרצליה הוא שם של עיר, שם של רחוב בחיפה, שם של בית חולים פרטי ושם של נשים בישראל. הוצעו מספר גישות לזיהוי מיקום גיאוגרפי על בסיס תוכן של מסר. קיימות טכניקות לזיהוי אוטומטי של שמות בתוך מלל. לאחר שאותרו מילים שזוהו כשמות, נדרש להפריד בין שמות המקומות לשמות אחרים. Web-a-Where, מערכת למציאת הקשרים גיאוגרפיים בדפי אינטרנט, הינה אחת העבודות החלוציות להתמודדות עם הבעיה (Amitay et al., 2004). TwitterTagger הינה מערכת שפותחה מאוחר יותר, ומתייגת מיקום גיאוגרפי של ציצי Twitter בהתבסס על השוואתם לבסיס נתוני המיקום של United States Geological Survey (USGS) (Paradesi, 2011).<sup>15</sup>



**איור 3:** תיאור סכמטי של תהליך הזיהוי שעל רשות עירונית לבצע במהלך חיפוש מסרים הנוגעים למערכת התחבורה במרחב העירוני שלה (Grant-Muller et al., 2014)

בתחום התחבורה ניתן להרחיב את העיקרון של שימוש בבסיסי נתונים של שמות מקומות, וליישם אותו גם על בסיס מידע הכוללים שמות ישויות תחבורתיות, כגון שמות/מספרי קווי תחבורה ציבורית, שמות חניונים וכד'. בסיסי מידע אלה מהווים נכס משמעותי בפיתוח מתודולוגיות לזיהוי אוטומטי של המרחב הגיאוגרפי אליו מתייחס המסר. דוגמא מעניינת ליישום שיטה זו נמצאת במאמרם של Bry et al. (2005), שבו מתוארת בניית מודל של נתונים מרחביים הכוללים גם מידע ספציפי הנוגע לשירותי תחבורה כגון רכבות.

ניתן לשער כי זיהוי אפקטיבי של המרחב הגיאוגרפי אליו מתייחס תוכן הנמצא במדיה חברתית צריך לשלב מספר טכניקות, בהתאם למאפייני המסר. באיור 3 נתון תיאור סכמטי של תהליך הזיהוי שעל רשות עירונית לבצע במהלך חיפוש מסרים הנוגעים למערכת התחבורה במרחב העירוני שלה.

### סיכום ומסקנות

עבודות מחקר ומקרי בוחן שעסקו בהפקת מידע תחבורתי ממדיה חברתית מעידים על יכולת לאתר תופעות, צרכים והעדפות של משתמשי הדרך. תרומתו של מידע זה משמעותית ביותר באזורים רבים אשר אינם מנוטרים באמצעים אחרים (כגון גלאים ומצלמות), ותרומה זו מאפשרת לרשויות לנקוט בפעולות אפקטיביות המתבססות על המידע החדש. לדוגמא, זיהוי כמעט מיידי של עצם קיום תאונה, ויותר מכך הבנת מאפייניה הספציפיים, מאפשרת טיפול יעיל וממוקד באירוע. במקרים אחרים, כגון בעת איסוף מידע על רמת שביעות הרצון של משתמשי התחבורה הציבורית, הפקת מידע ממדיה חברתית מציבה חלופה זולה יחסית לאיסוף מידע מסוג זה, וכזו הנותנת ביטוי להיבטי השירות שמשמשי הדרך בחרו להבליט.

בו בזמן, המורכבות של "הפרדת המוץ מהתבן", קרי זיקוק המידע הרלוונטי והמשמעותי, עדיין רבה וסוגיות רבות טרם נפתרו. טכניקות כריית המלל בכלל, ואלה המיועדות לתחום התחבורה בלבד, טרם הגיעו לרמת בשלות המבטיחה אחוזי הצלחה גבוהים הן במונחים של דיוק והן במונחים של איחזור. העובדה כי המידע המופק עלול להיות מוטא, ולא דווקא מייצג את כלל משתמשי הדרך, מעצים את הצורך לפתח שיטות להיתוך מידע זה עם מקורות מידע מסורתיים, קרי סקרים לסוגיהם השונים.

ניתן לסכם ולומר כי מקומה של המדיה החברתית כערוץ תקשורת הולך ומתעצם בחברה המודרנית. לכן, בעוד האתגרים בהפיכתה לזירה אפקטיבית להפקת והפצת מידע תחבורתי עדיין רבים, קיימות ראיות מבטיחות כי המשך המחקר והפיתוח בתחום זה יאפשר את מיצוי הפוטנציאל הרב הטמון בה.

### הערות

1. <http://www.gartner.com/it-glossary/social-media>
2. <http://www.tapuz.co.il/forums2008/forumpage.aspx?forumid=1630&r=1>
3. [/http://www.tripadvisor.co.il](http://www.tripadvisor.co.il)
4. [/https://www.facebook.com](https://www.facebook.com)

- .5 /http://www.blogger.co.il
- .6 /https://twitter.com
- .7 /http://en.reddit.com
- .8 /https://www.youtube.com
- .9 https://twitter.com/itayg1/status/565597769994625024
- .10 https://twitter.com/pinhaswolf/status/464295996797616129
- .11 https://www.facebook.com/amir.aharoni posts/101541255 51240076
- .12 https://twitter.com/elladoron/status/356148398858059778
- .13 /https://tahbura.wordpress.com/project\_sardine
- .14 https://twitter.com/attilus/status/567060642398343169
- .15 /http://geonames.usgs.gov

### מקורות

- Amitay, E., Har'El, N., Sivan, R., Soffer, A., (2004), "Web-a-Where: "Geotagging Web Content", SIGIR '04 Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 273-280
- Bry, F., Lorenz, B., Ohlbach, H.J., Rosner, M., (2005), "A Geospatial World Model for the Semantic Web", Principles and Practice of Semantic Web Reasoning, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Berlin, Heidelberg, 3703, pp. 145–159.
- Carrasco J A, Hogan B, Wellman B, Miller E J, (2008),"Collecting social network data to study social activity-travel behavior: an egocentric approach" *Environment and Planning B: Planning and Design* 35(6) 961 – 980.
- Cho S., Kang J.Y., Yasar, A. Luk Knapen L., Bellemans T., Janssens D., Wets G., Hwang C., (2013), "An Activity-based Carpooling Microsimulation using Ontology", *Procedia Computer Science*, 19, pp. 48–55
- Collins, C., Hasan, S., Ukkusuri, S.V., (2013), "A Novel Transit Rider Satisfaction Metric: Rider Sentiments Measured from Online Social Media Data", *Journal of Public Transportation*, 16, (2), pp. 21-45
- Corley, C., Cook, D., Mikler, A., Singh, K., (2010), "Text and structural data mining of influenza mentions in web and social media", *International Journal of Environmental Research and Public Health*, 7, (2), pp. 596–615
- Gal-Tzur, A., Grant-Muller, S. M., Kuflik T., Minkov, E., Nocera, S., Soor I., (2014a), "The Potential of Social Media in Delivering Transport Policy Goals", *Transport Policy*, Vol 32, pp 115-123



Gal-Tzur, A., Grant-Muller, S. M., Minkov, E., Nocera, S., (2014b), The Impact of Social Media Usage on Transport Policy: Issues, Challenges and Recommendations", *Procedia - Social and Behavioral Science*, 2014, Vol 111, pp 937-946

Grishman, R., Huttunen, S., Yangarber, R., (2002), "Information extraction for enhanced access to disease outbreak reports", *Journal of Biomedical Informatics*, 35, (4), pp. 236–246

Grant-Muller S.M., Gal-Tzur A., Minkov E., Nocera S., Kuffik T., and Shoor I., (2014), "Enhancing Transport Data Collection through Social Media Sources: Methods, Challenges and Opportunities for Textual Data", IET Intelligent Transport Systems. doi: 10.1049/iet-its.2013.0214

Kaplan A.M., Haenlein M. (2010), "Users of the world, unite! The challenges and opportunities of SocialMedia", *Business Horizons*, Vol. 53 Issue 1, PP. 59-68

Kushal, D., Lawrence, S., Pennock, D. M., (2003), "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", *Proceedings of the 12th international conference on World Wide Web*, ACM, pp. 519-528.

Lanagan, J., and Smeaton, A. F., (2011)"Using Twitter to Detect and Tag Important Events in Live Sports", in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM)

Liu X., Lang B., Yu W., Lou J. Huang L., (2011), "AUDR: An Advanced Unstructured Data Repository", *Pervasive Computing and Applications (ICPCA)*, 2011 6th International Conference, Conference Publication PP. 462-469

Mai, E., Hranac, R.. (2013), "Twitter Interactions as a Data Source for Transportation Incidents", *TRB 92nd Annual Meeting Compendium of Papers*.

Niaraki A.S., Kim, K., (2009), "Ontology based personalized route planning system using a multi-criteria decision making approach", *Expert Systems with Applications*, 2009, 36, pp. 2250–2259.

Pak, A., Paroubek, P., (2010), "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", *Computer*, 10, pp. 1320-1326.

Nocera, S., (2011), "The key role of quality assessment in public transport policy", *Traffic Engineering & Control*, 52, (9), pp. 394-398.

Pang, B., Lee, L., (2008), "Opinion Mining and Sentiment Analysis", *Foundations and Trends Information Retrieval*, 2, (1-2), pp. 1-135.

Paradesi, S., (2011), "Geotagging Tweets Using Their Content", Proceedings of the Twenty-Fourth. International Florida Artificial Intelligence Research Society Conference, pp. 335-356.

Tumasjan, A., Sprenger, T.O., Sandner, P. G., Welpe, I. M., (2010), "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Pender B., Currie G., Delbosc A., Shiwakoti N., (2014), "Social Media Use in Unplanned Passenger Rail Disruptions - An International Study", TRB 93rd Annual Meeting.

Schulz, A., Ristoski, P. and Paulheim, H., (2013), "I see a car crash: Real-time detection of small scale incidents in microblogs", In P. Cimiano, M. Fern`andez, V. Lopez, S. Schlobach, and J. V`olker, editors, 'The Semantic Web: ESWC 2013 Satellite Events,' number 7955 in Lecture Notes in Computer Science, pages 22–33. Springer Berlin Heidelberg

Schweitzer, L., (2012), "How Are We Doing? Opinion Mining Customer Sentiment in US Transit Agencies and Airlines Via Twitter", Presented at the 91th Annual Meeting of the Transportation Research Board, Washington, D.C.

Sterne J., 2010, "Social media metrics: How to measure and optimize your marketing investment", books.google.com

Steiger E., Ellersiek T., Zipf A., (2014), "Explorative public transport flow analysis from uncertain social media data", GeoCrowd '14 Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, PP. 1-7

Steinwart I., Christmann A., (2008), "Support Vector Machines", Springer, New York.

Wang J., Ding, Z., Jiang, C., (2005), 'An Ontology-based Public Transport Query System', Proceedings of the First International Conference on semantics and Grid', SKG