

## ניתוח השיח במאמרי כנסי צ'ייס לחקר חדשנות וטכנולוגיות-למידה בגישה מונעת-נתונים

ניצה גרי	יורם עשת-אלקלעי	ורד זילבר-ורוד
האוניברסיטה הפתוחה	האוניברסיטה הפתוחה	האוניברסיטה הפתוחה
<a href="mailto:nitzage@openu.ac.il">nitzage@openu.ac.il</a>	<a href="mailto:yorames@openu.ac.il">yorames@openu.ac.il</a>	<a href="mailto:vereds@openu.ac.il">vereds@openu.ac.il</a>

### Analyzing the Discourse of the Articles of Chais Conferences for the Study of Innovation and Learning Technologies via a Data-driven Approach

Vered Silber-Varod      Yoram Eshet-Alkalai      Nitza Geri  
The Open University of Israel      The Open University of Israel      The Open University of Israel

#### Abstract

One of the challenges facing innovative technologies researchers is developing methodologies and measures for improving the description and prediction of assimilation and diffusion of technologies in organizations. Data-driven approach is one of the contemporary methodologies in this field. In recent decades, this approach has evolved into an independent field of knowledge, known as the Corpus Linguistics – a field based on sophisticated algorithms for natural language processing, which are used for investigating structural and contextual linguistic characteristics of texts. In the current study, we applied this approach for analyzing a database of articles published in the proceedings of Chais Conferences for Innovation and Learning Technologies during 2006-2014. The purpose of the study was to identify the specific terms (words and phrases) used in the research of learning technologies. An exhaustive process of analyzing all the words in the corpus unexpectedly revealed that the terms most commonly used in the field (e.g., pupil, teacher, student) and the phrase most commonly used (face-to-face) actually belong to a pedagogical rather than a technological genre. The findings pointed out the potential and limitations of a data-driven approach in identifying changes over time in the field of learning technologies.

**Keywords:** quantitative text analysis, mapping topics in the domain of learning technologies, word frequency, keyword generation methods, culturomics.

#### תקציר

אחד האתגרים העומדים בפני חוקרי טכנולוגיות חדשניות הוא פיתוח מתודולוגיות ומדדים שישפרו את התיאור והחיזוי של הטמעת והתפשטות טכנולוגיות בארגונים. גישה מונעת-נתונים (data-driven) היא בין המתודולוגיות העכשוויות בתחום זה. בעשורים האחרונים התפתחה גישה זו לכדי תחום-דעת עצמאי, המכונה 'בלשנות הקורפוס' (Corpus Linguistics) – תחום המבוסס על שימוש באלגוריתמים משוכללים לעיבוד שפה טבעית (natural language processing) – אלגוריתמים המשמשים לחקירת מאפיינים לשוניים, מבניים ותוכניים, של טקסטים. במחקר הנוכחי נעשה שימוש בעיבוד כזה על בסיס נתונים של מאמרי כנסי צ'ייס לחקר חדשנות וטכנולוגיות למידה בשנים 2006-2014, על מנת לזהות את המונחים הייחודיים (מילים וצירופי מילים) למחקרים בתחום טכנולוגיות הלמידה. מניתוח הממצאים בתהליך ממצה

(exhaustive) של כל מילות התחום, נמצא במפתיע כי המונחים הנפוצים ביותר במחקרים בתחום (למשל: 'תלמיד', 'מורה', 'סטודנט') וגם צירוף המילים הנפוץ ביותר (פנים-אל-פנים) הם דווקא מתחום הפדגוגיה ולא מתחום הטכנולוגיה. הממצאים הצביעו על הפוטנציאל ועל המגבלות של הגישה מונעת-הנתונים לזיהוי תמורות לאורך זמן בתחום טכנולוגיות הלמידה.

**מילות מפתח:** ניתוח טקסט כמותי, מיפוי נושאים בתחום טכנולוגיות למידה, שכיחות מילים, שיטות להפקת ביטויי מפתח, תרבוטרמיקה (culturomics).

## הקדמה

בעולם המתאפיין בשינויים טכנולוגיים מהירים, עומדים חוקרי טכנולוגיות למידה מול מצב פרדוקסלי, בו הם בוחנים כיצד מתמודדים אנשי הוראה ותלמידים עם סביבות למידה המשתנות חדשות לבקרים (Cuban 1999). פרדוקס זה בא לידי ביטוי בתחלופה בקצב הולך וגובר, של טכנולוגיות למידה חדשניות במערכת החינוך – דבר המקשה על הטמעה מיטבית ועל ביצוע מחקרים מעמיקים הבוחנים את אפקטיביות השימוש בטכנולוגיות אלה. מטרת המחקר הנוכחי היא לבחון את התמורות המתחוללות בטכנולוגיות למידה לאורך זמן, כפי שהן משתקפות בטקסטים של מאמרים שהתפרסמו במהלך העשור האחרון (בשנים 2006-2014) במסגרת כנסי ציירי לחקר חדשנות וטכנולוגיות למידה, שמארגן המרכז לחקר חדשנות בטכנולוגיות למידה באוניברסיטה הפתוחה. התמות המרכזיות של הכנס הן בחינה אמפירית של מודלים של למידה בסביבות טכנולוגיות; חקר היבטים קוגניטיביים של למידה בסביבות טכנולוגיות; בחינת אפקטיביות ההוראה באמצעות כלים ויישומים מונחי-תוכן; חקר האינטראקציה בין לומדים למחשב; טכנולוגיה כמגשרת ומצמצמת מרחקים ופערים; חקר של שילוב טכנולוגיות במערכת החינוך ובמוסדות להשכלה גבוהה בישראל; בחינת פרדיגמות חדשניות בהדרכה ולמידה בארגונים ובתעשייה; טכנולוגיות למידה לקידום אנשים עם צרכים מיוחדים; טכנולוגיות למידה פורצות דרך; מודלים חדשניים בלמידה מקוונת; משאבי למידה פתוחים (Open educational resources), Massive Open Online Courses (MOOCs), למידה לאורך החיים (Lifelong learning); טכנולוגיות למידה וצמצום פערים חברתיים; משחקים דיגיטליים כסביבות למידה; מיקור המונים (crowdsourcing); אינטראקציה ברשתות חברתיות; היבטים כלכליים, ארגוניים, חוקיים ואתיים של שילוב טכנולוגיות למידה; מדידת הערך המוסף של טכנולוגיות חדשות ושל ניהול מידע בארגונים ובמוסדות; התנגדויות לאימוץ טכנולוגיות חדשות בארגונים ובמוסדות.

במחקר הנוכחי מודגם שימוש חדשני בגישה מונעת-נתונים (data-driven approach) לניתוח מגמות בשיח האקדמי והמחקרי בתחום טכנולוגיות הלמידה.

## ניתוח שיח בגישה מונעת-נתונים

בגישה מונעת-נתונים החוקרים מתבססים על נתונים אמפיריים טבעיים הנובעים ישירות מקורפוס הנתונים הנחקר. מכיוון שהקורפוס מורכב מטקסטים, התפתח תחום הדעת הבלשני הקרוי "בלשנות הקורפוס" (Sinclair 1991; 2004), שבו בנוסף לזיהוי המבנים במערכת הלשונית, ממפים את אופן השימוש בשפה, בבחינת "האמינו לטקסט", ככותרת ספרו של סינקלייר (Sinclair, 2004) *Trust the text: Language, corpus and discourse*. החוקרים נעזרים בקורפוס כדי לתאר מאפיינים בלשון מסוימת או סוגה מסוימת. הקורפוס הנבחר חייב להיות מייצג (representative) מההיבט של הנושא הנחקר (לבנת, 2010).

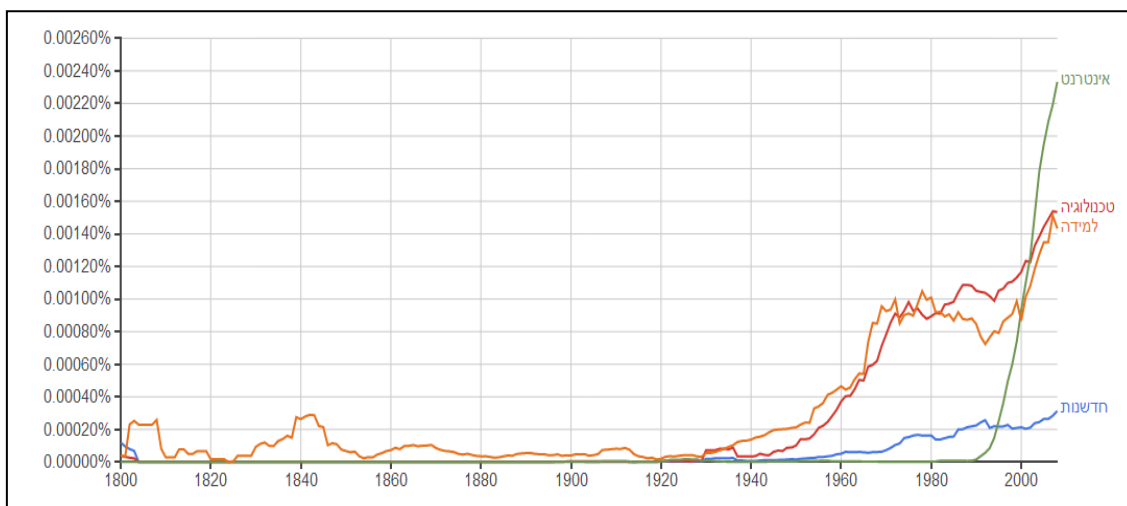
לשם המחשת הכוח האנליטי שיש בגישה מונעת קורפוס, יצרו Michel et al. (2011) מערך נתונים המבוסס על 5,195,769 ספרים דיגיטליים מהשנים 1800-2000, מערך שהכיל כ-4% בקירוב מכל הספרים שפורסמו אי פעם (באנגלית) (Michel et al., p.176). הם השתמשו בניתוח בשיטת כריית נתונים המכונה n-gram, בכדי למדוד מגמות תרבות, כפי שהן באות לידי ביטוי בבחירות הלשוניות ובטקסטים עצמם. דוגמה מעניינת לשימוש מחקרי בגישה מונעת-הנתונים לניתוח שיח ניתן למצוא אצל Johnston (2008), המציגה מחקר שבחן, באמצעות בדיקת השיום והמינוח של המאמרים בתחום, כיצד חוקרים מגדירים תהליכים ארוכי-טווח כגון הזדקנות מול הזדקקות לצרכים מיוחדים (Johnson, 2008, p.7).

האצת השימוש בגישת בלשנות הקורפוס נבעה מפיתוח אלגוריתמים משוכללים לעיבוד שפה טבעית (natural language processing). השימוש בשיטות סטטיסטיות מתקדמות לניתוח נתונים של בסיסי נתונים טקסטואליים רחבי היקף, מאפשר לזהות דפוסי שיוני ולקודד תכונות מהותיות הבאות לידי ביטוי בטקסטים.

אחד הענפים שהתפצל מתחום בלשנות הקורפוס נקרא תרבוּמטריקה (culturomics) (Michel et al., 2011), ועוסק בחקירה כמותית של מערכות טקסט ושפה מדוברת, מאסיביות ודיגיטליות, לבחינת דפוסים תרבותיים בדיסציפלינות שונות (Webber, 2001; Bohannon, 2011). התרבוּמטריקה גם מאפשרת להתחקות אחר מגמות שינוי ולזהות את המינוח הייחודי (מילים וצירופי מילים) בתחום מסוים על ידי ניתוח הפרסומים באותו תחום-דעת (Soper & Turel 2012a, 2012b; Soper, Turel, & Geri 2014).

יישום נוסף של אנליזות כמותיות של טקסטים, אשר תופס תאוצה בשנים האחרונות, הוא הערכה אוטומטית למחצה של מבחנים (Bennet & Ben-Simon, 2005). יישום זה פותח במרכז הארצי לבחינות והערכה (שפרן ובן-סימון, 2014). בעקבות התפתחות בלשנות הקורפוס, פותחו כלים אוטומטיים וחצי-אוטומטיים, למשל, Matrix (Rayson 2003), ו-Antconc (Anthony, 2011), המקלים ומאיצים את הניתוח. שימוש בכלים כאלה מאפשר להתבונן בגוף הידע (קורפוס) כולו ולא רק בתכונות מסוימות שלו. בעידן ה-Big Data, חוקרים מדיסציפלינות שונות אימצו שיטות חישוביות כאלה ואחרות, המתבססות על רמת השכיחות של מילים ויחסים בין רמות שכיחות שונות של מילים לפיתוח תחזיות בתחומי הכלכלה והפוליטיקה – למשל באתרי חדשות מקוונים (Radinsky, Agichtein, Gabrilovich, & Markovitch, 2011; Radinsky & Horvitz, 2013).

על מנת לבחון את הפוטנציאל של ניתוח כזה לזיהוי מגמות שינוי לאורך זמן בתחום טכנולוגיות הלמידה, ביצענו בדיקה בגוגל-ספרים (Google Books) בעברית, שהוא בסיס נתונים פתוח ודיאכרוני (מכיל נתונים הנפרשים על טווח של שנים רבות) של ספרים. תרשים 1 מראה כיצד המילים "למידה", "אינטרנט" ו"טכנולוגיה", ו"חדשנות" הופיעו (ונעלמו) בספרות במהלך 200 השנים האחרונות, ואת יחסי השכיחות ביניהן. ואכן, ניתוח טקסטואלי יכול ללמד לא רק על דפוסי שינוי בזמן של מילה בודדת, אלא גם על המועד שבו המילה הופיעה לראשונה והמועד שבו נעלמה. מהגרף עולה שהמילה "למידה" הייתה בשימוש כבר בתחילת המאה ה-19 (וסביר להניח שגם קודם לכן). גם המילה "טכנולוגיה" הייתה קיימת במאה ה-19, אך השימוש בה גבר משמעותית החל משנות השלושים של המאה ה-20, במקביל להופעת הטלוויזיה בצורתה הגולמית בסוף שנות העשרים. המילה חדשנות מתחילה להופיע בשנות החמישים, ואילו המילה 'אינטרנט' פורצת, פשוטו כמשמעו, החל משנות התשעים של המאה ה-20. כמו כן, מצביע תרשים 1 על ההופעה הבו-זמנית של שתי המילים "למידה" ו"טכנולוגיה".



**תרשים 1. שכיחות השימוש בארבעה מונחים בתחום טכנולוגיות למידה – אינטרנט, טכנולוגיה, למידה וחדשנות – על פני שתי מאות, כפי שאוחזר מבסיס הנתונים של הספרים בעברית בפלטפורמת Google Books Ngram Viewer**

### שאלת המחקר

מהם דפוסי השינוי לאורך זמן בדגשים המחקריים של טכנולוגיות למידה, כפי שהם משתקפים במילות המפתח ובטקסט המלא של קורפוס המאמרים שפורסמו בכנסי צ'ייס בשנים 2006-2014.

## שיטת המחקר

המחקר נקט בגישה מונעת-נתונים שתוארה לעיל, המבוססת על ניתוח כמותי של טקסטים (Bauer, 2000). מכיוון שפרמטר ההקשר של הופעת המונחים הינו ידוע ונתון (מאמרי כנסי צייס מהשנים 2006-2014), לא נערך קידוד תוכן איכותני, אלא אינפורמטיבי: לכל מילה הוצמדו שנת הופעתה ושייכה למאמר.

## קורפוס הנתונים

כקורפוס הנתונים למחקר שימשו 1,408 **מילות המפתח**, אשר נכתבו על ידי מחבריהם של 553 המאמרים והפוסטרים של כנסי צייס בשנים 2006-2014 (383 בעברית ו-170 באנגלית) וכן **הטקסט המלא** של המאמרים עצמם.

ניתוח המאמרים עצמם בוצע בשתי שיטות. האחת היא הפקת **רשימות ברמת המילה הבודדת**, כפי שיוסבר להלן, והשנייה היא בדיקת **רצף המילים במאמר**. בדיקה של רצפי המילים (שיטה הקרויה n-gram, ומשמשת לחקר תרבומטריקה) מכמתת שכיחות של כל רצף מילים, לאו דווקא של ביטוי שכיח ומוכר כמו "בית ספר", אלא של כל שתי מילים ויותר המופיעות בטקסט זו אחר זו (כמו, למשל, הרצף "נמצא הבדל מובהק", שמופיע בבסיס הנתונים של כנסי צייס 65 פעמים ב-39 מאמרים שונים).

## תהליך הפקת רשימות המילים

אחת השיטות הנפוצות כיום לעיבוד מידע מטקסטים, היא הכנת רשימות שכיחות ברמת המילה הבודדת. ואולם, שיטה זו לניתוח טקסט טבעי היא בעלת מגבלות רבות. למשל, הסתמכות על ספירת מילים כתובות בלבד, כלומר כל רצף אותיות-גראפמות בין רווחים, הנה מטעה משום ש"הלמידה" ו"בלמידה", למשל, ייספרו כשתי מילים שונות. לפיכך, היחידה לניתוח שבה מעבדים שפה טבעית, כלומר טקסטים, היא "לְמָה" (lemma), שהיא צורת הבסיס של המילה. הכוונה היא לתהליך איחוד של מילים בעלות נטייה שונה, אך משמעות זהה ומאותה קטגוריה לקסיקלית. למשל, הִלְמָה "תלמיד" היא צורת הבסיס של המילים: תלמידה, תלמידים, תלמידות, תלמידי, תלמידותיו, תלמידו, וכדומה.

את עיבוד המילים מצורתן הכתובה לצורת הבסיס שלהן ביצענו במחקר הנוכחי בסיוע המרכז הארצי לבחינות ולהערכה (National Institute for Testing and Evaluation), שברשותו כלים מתקדמים לניתוח מורפולוגי של השפה העברית (לסקירת הכלים השונים שמרכיבים את מערכת העיבוד ראו באתר **פרייקט השפה העברית** <https://hlp.nite.org.il>, וכן: Ben-Simon & Cohen, 2011; Ben-Simon & Safran 2012).

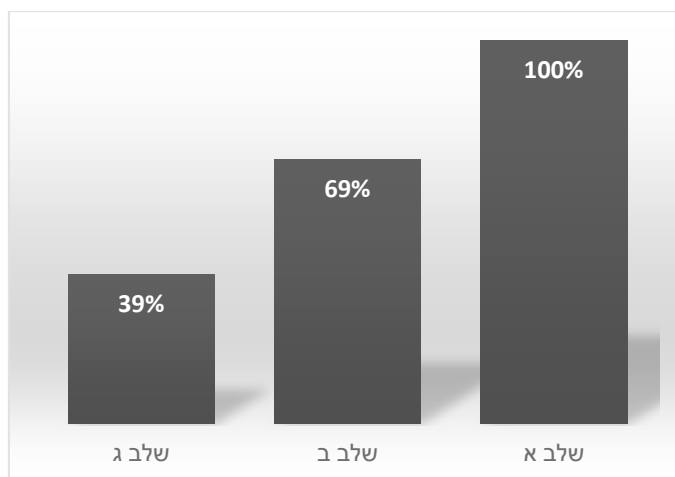
תהליך הכנת רשימות המילים כלל ניקוי המאמרים מכותרות ראשיות, כותרות משנה, שמות מחברים, רשימות מקורות, טבלאות ותרשימים, והשאררת גוף הטקסט בלבד. לאחר מכן התבצע תהליך ה"למטיזציה" (lemmatization), שהוסבר לעיל. במקור, נערך ניתוח טקסטואלי נפרד למאמרים בעברית ולאנגלית ולאחר מכן אוחדו הממצאים ידנית, בתהליך המתואר בטבלה 1.

**טבלה 1. בסיס הנתונים של מאמרי צ'ייס 2006-2014**

בסיס הנתונים הכולל	אנגלית	עברית	
<b>שלב א:</b> בסיס הנתונים מבוסס על מאמרי כנס צ'ייס שהתפרסמו בספרי הכנס בשנים 2006-2014 (ראו נספח א לרשימת ספרי הכנס ומראי המקום).			
730,075	תמניות 249,769 תבניות 10,886	תמניות* 480,306 תבניות** 42,008	
<b>שלב ב:</b> יצירת רשימות מילים מהטקסט השוטף. בסיס הנתונים בעברית עבר תהליך למטיזציה. ומבסיס הנתונים באנגלית הוסרו כל מילות הפונקציה.			
507,178	תמניות 124,237 תבניות 10,431	תמניות 382,941 תבניות 12,310	
<b>שלב ג:</b> צמצום בסיס הנתונים לשמות: שמות עצם, שמות תואר ושמות פרטיים (למשל: למידה, טכנולוגיה, חינוך, שיתופי, פייסבוק), וצמצום הרשימה ל-549 המילים בעברית שנמצאו בשכיחות של 100 הופעות ומעלה במהלך השנים. שילוב המילים באנגלית אל תוך רשימת המילים בעברית בהתאמה ידנית.			
287,118	תמניות 60,499 תבניות 506	תמניות 226,619 תבניות 549	

\* המונח "מילה תמנית" (מלשון 'למנות', לספור) מתייחס למספר ההופעות הכולל של מילה בודדת בקורפוס.  
\*\* המונח "מילה תבנית" מתייחס לכמות המילים השונות שמופיעות בקורפוס.

תרשים 2 ממחיש את תהליך צמצום רשימת המילים כתוצאה מתהליך הלמטיזציה וצמצום מילים בשכיחות נמוכה במיוחד. יחס המיצוי (recall ratio) מתחילת התהליך ועד סופו הוא 39%.



**תרשים 2. המחשה של תהליך מיצוי רשימת המילים הסופית לניתוח המאמרים**

**עיבוד הנתונים וניתוחם**

על הקורפוס שנוצר בשלב ג לעיל, ערכנו את הבדיקות הבאות: המילים הבסיסיות של התחום; המילים הבולטות בכל שנה; חישוב ממד ההשתנות בזמן.

ראשית, קבענו את המילים הבסיסיות של התחום שהיו בעלות שכיחות גלובלית מעל 2,000 הופעות במהלך השנים. בשיטה זאת ציפינו לאתר את המילים המופיעות מספר רב של פעמים מדי שנה.

מילה בסיסית הוגדרה כמילה שהופיעה בכל תשע השנים מספר פעמים שהוא במרחק לכל היותר סטיית תקן אחת מממוצע השכיחות היחסית שלה.

חישוב **בולטות המילה** בשנה מסוימת לעומת כל השנים שלפניה (למעט שנת 2006) נעשה באמצעות חישוב סטטיסטי המכונה log-likelihood (Rayson & Garside, 2000), המאפשר להשוות בין שכיחויות יחסיות של מילים מקורפורה שונים.

חישוב **ממד ההשתנות עם הזמן** נערך באמצעות השוואת נתוני ה-log-likelihood בכל אחת מהשנים 2014-2011 לשנים הקודמות להן.

## ממצאים

בסעיף זה יוצגו הממצאים, תוך הבחנה בין המגמות שהצטיירו מניתוח מילות המפתח (keywords) לאלה שנמצאו בניתוח שכיחויות המילים הבודדות ורצפי המילים.

### מילות מפתח

מילת המפתח השכיחה ביותר הייתה "למידה שיתופית" (הופיעה 40 פעמים כמילת מפתח). אחריה מופיעות "למידה מרחוק" ו"למידה מתוקשבת" ב-25 ו-23 מאמרים בהתאמה. "השכלה גבוהה", "ויקי", "תקשוב", "רשת חברתית", "הכשרת עובדי הוראה", "פייסבוק" ו"למידה" עם שכיחות הופעות בין 18 ל-11 מאמרים בלבד. כל יתר 1,398 מילות המפתח הופיעו בעשרה מאמרים או פחות. רוב הפריטים (1,061 מילות מפתח) מופיעים פעם בודדת (singletons). בהתאמה, התפלגות השכיחויות עבור 1,408 מילות המפתח בשנים 2006-2014, המוצגת בתרשים 3, מאופיינת בתופעת "הזנב הארוך" ("התפלגות זיפף" או Zipf's law), והיא התפלגות שאופיינית לקורפוס טקסטואלי טבעי (Rayson, 2003, p. 36).



תרשים 3. התפלגות השכיחויות של מילות המפתח 2014-2006

### רצפי מילים

בבדיקה זאת נבדקו 430,161 רצפים. הרצף השכיח ביותר שמופיע בבסיס הנתונים הוא הביטוי "פנים אל פנים" (356 הופעות ב-113 מאמרים). 20 הרצפים השכיחים ביותר אחריו הם כמעט כולם מסוג המעיד על הכתיבה המדעית (למשל, "נמצא הבדל מובהק") ולא על התוכן. יחד עם זאת, הרצף במקום ה-14 הוא "מורים בבתי ספר" (95 הופעות ב-11 מאמרים). ובמקום ה-20 – "חשיבה מסדר גבוה" שמופיע 39 פעמים ב-17 מאמרים.

### המילים הבסיסיות במאמרי צייס – ה-DNA של התחום

בניגוד לשני הסעיפים הקודמים שעסקו במילות מפתח וברצפים טבעיים בטקסט, בחלק זה נציג ממצאים המתאייסות לרשימות מילים בודדות.

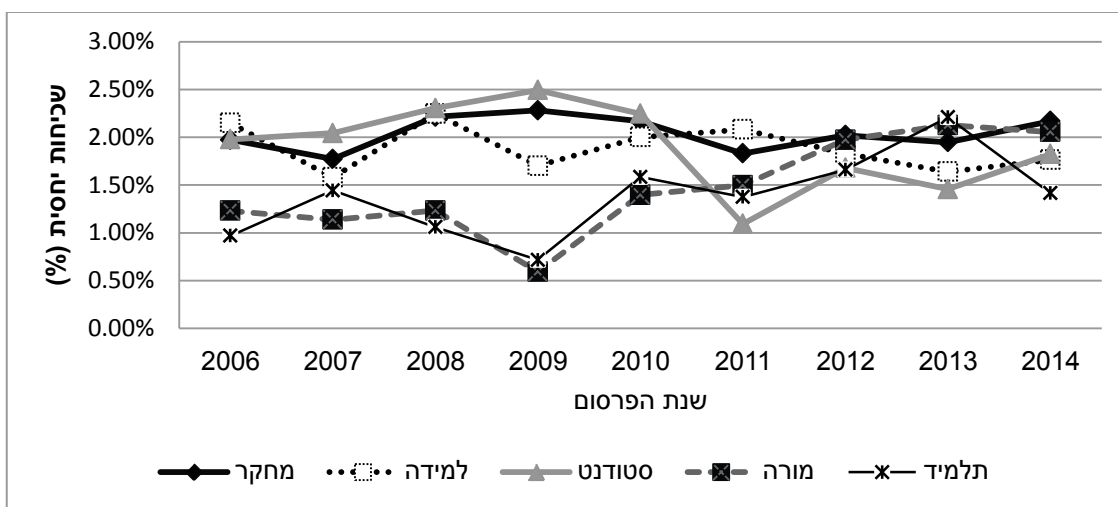
בטבלה 2 מוצגות חמש המילים השכיחות ביותר שנמצאו בניתוח הגלובלי. המילה 'מחקר' היא המילה השכיחה ביותר והיא יוצאת דופן בכך שהיא משקפת את העובדה שמדובר בקורפוס של

מאמרים מחקריים, לעומת ארבע המילים האחרות המשקפות את מושאי המחקר). בטבלה מוצגות צורות הבסיס של המילים.

**טבלה 2. חמש המילים הבסיסיות במחקר בתחום טכנולוגיות למידה  
כפי שמתקף במאמרי צ'ינס 2014-2006**

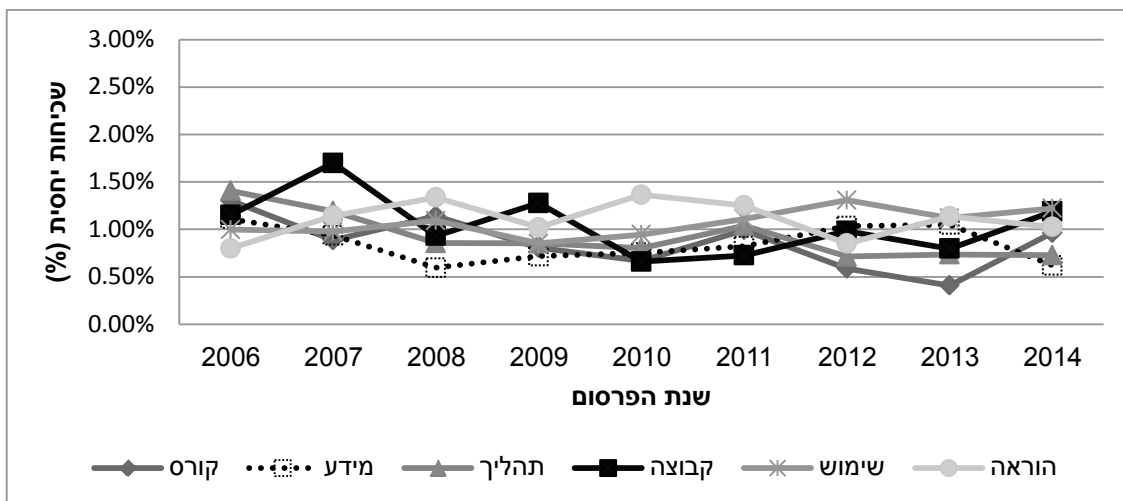
מחקר	2006	2007	2008	2009	2010	2011	2012	2013	2014	ממוצע שכיחות יחסית	סטיית תקן
מחקר	1.97%	1.77%	2.21%	2.28%	2.17%	1.83%	2.02%	1.95%	2.17%	2.04%	0.18%
למידה	2.15%	1.58%	2.25%	1.70%	2.00%	2.08%	1.83%	1.64%	1.76%	1.89%	0.24%
סטודנט	1.98%	2.04%	2.30%	2.49%	2.25%	1.09%	1.68%	1.46%	1.82%	1.90%	0.44%
מורה	1.23%	1.14%	1.23%	0.59%	1.40%	1.50%	1.97%	2.13%	2.06%	1.47%	0.50%
תלמיד	0.97%	1.44%	1.06%	0.72%	1.58%	1.37%	1.66%	2.21%	1.41%	1.38%	0.44%

תרשים 4 ממחיש את הנתונים המוצגים בטבלה 2. מתרשים זה וגם מתרשימים 5-6, ניתן להבחין בחוסר התנודתיות שבהופעת המילים ומכך שהן פחות או יותר מופיעות באותה רמת שכיחות לאורך תשע השנים שנחקרו, אם כי חשוב לציין שכל אחת מהמילים השכיחות ביותר נמצאה לפחות בשנה מסוימת חורגת בכמות ההופעות שלה מסטיית התקן, כלפי מטה או כלפי מעלה.

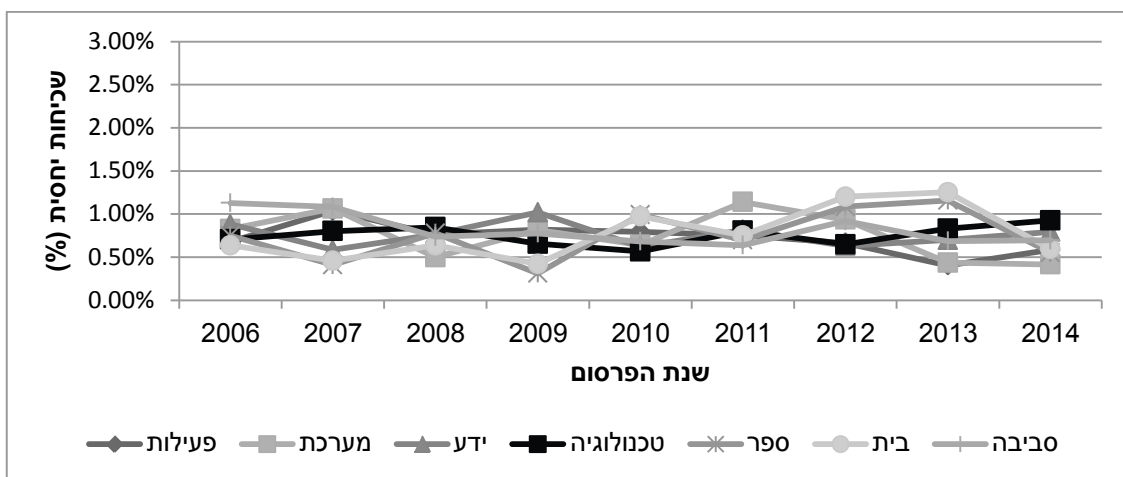


**תרשים 4. חמש המילים הבסיסיות במחקר בתחום טכנולוגיות למידה  
כפי שמתקף במאמרי צ'ינס 2014-2006**

המילים ששכיחותן נמוכה יותר, אך עדיין בולטות ביותר הן: קורס, קבוצה, תהליך, הוראה, סביבה, מידע, שימוש, בית, ספר, טכנולוגיה, ידע, פעילות ומערכת (תרשים 5 ותרשים 6).



תרשים 5. שש המילים הבסיסיות ששכיחותן היחסית נעה בין 0.4% ל-1.7%



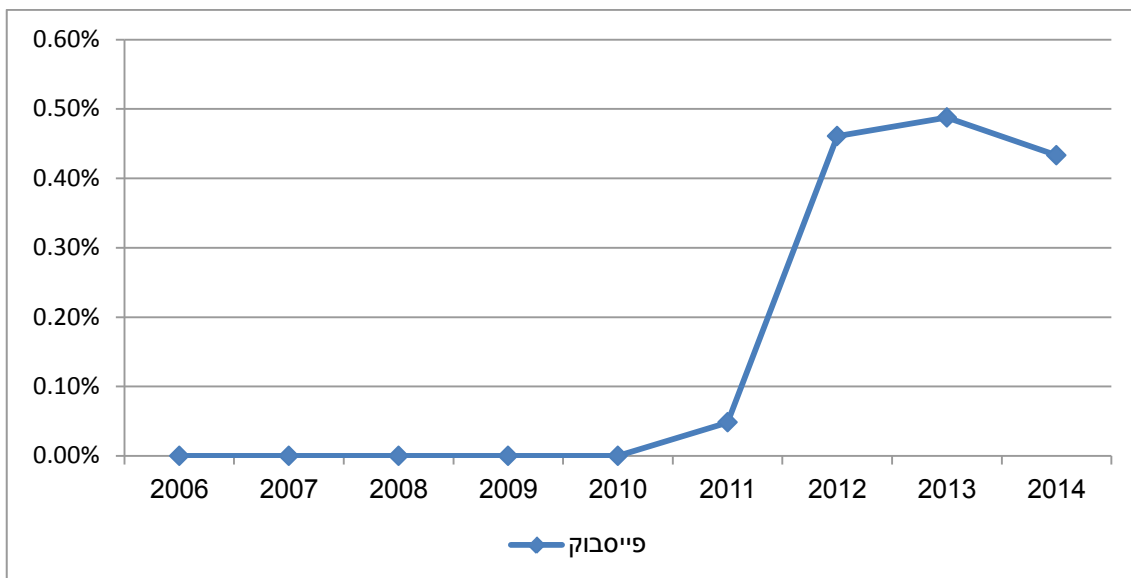
תרשים 6. שבע המילים הבסיסיות ששכיחותן נעה בין 0.3-1.2 אחוזים

#### מילים בעלות שכיחות בולטת במיוחד

מילים בולטות הן מילים שבשנה מסוימת הופיעו בשכיחות יחסית גבוהה באופן ניכר בהשוואה לשנים הקודמות. באותה מידה, מילה שלא הופיעה בשנים הקודמות ומופיעה לראשונה בשנה הנבדקת, אפילו פעמים ספורות בלבד, תקבל גם היא ערך גבוה בחישובי ה- $\log\text{-likelihood}$  (Rayson, 2008). בשל מגבלת ההיקף של המאמר, לא ניתן לפרוס בצורה ממצה את הנתונים שהעלו חישובי הבולטות ולפיקך בחרנו להציג כאן שני מקרים מעניינים שעלו מהם (תרשימים 7 ו-8).

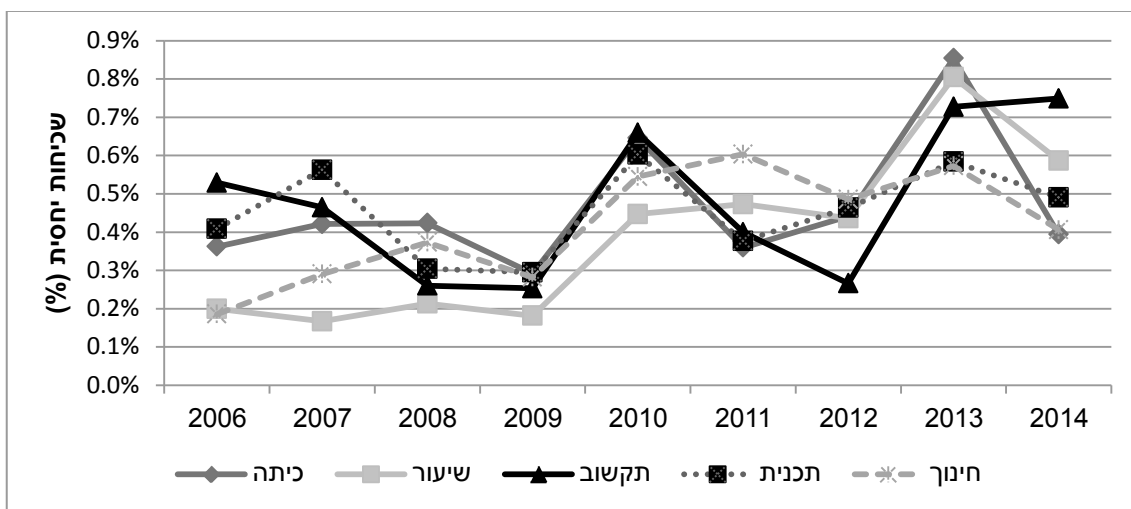
תרשים 7 מציג את השינוי לאורך זמן בהופעת המילה 'פייסבוק'. מהנתונים עולה כי בשנת 2011, מילה זו בולטת משמעותית, זאת לאחר שבשנים 2006-2010 היא לא הוזכרה כלל במאמרי צייס. החל משנת 2011 היא ממשיכה לבלוט משמעותית בכל שנה. את היעדרה של המילה פייסבוק בשנים 2006-2010 ואת הפצעת המילה מ-2011 ואילך (על אף שפייסבוק נפתחה לשימוש הציבור כבר ב-2006), אפשר להסביר בכך שרק בשנת 2009 נפתחה האפשרות להגדיר בפייסבוק שפות שאינן אנגלית (כולל עברית) לשימוש בפרופיל האישי. אנתנו סוברים כי האפשרות להגדרת שפות שונות הרחיבה משמעותית את השימוש ברשת חברתית זו, בעיקר בקרב בני הנוער, דבר שהוביל למחקרים על פייסבוק גם בתחום של טכנולוגיות למידה.





**תרשים 7. התפתחות השימוש במילה "פייסבוק" לאורך השנים**

תרשים 8 מציג את חמש המילים הבולטות, שהן בעלות השכיחות ביותר כלפי מעלה בשנת 2013: כיתה, תקשוב, תכנית, חינוך ושיעור. גם כאן, ניתן להסביר את הממצאים שהתקבלו מהניתוח הכמותי כמשקפים את המחקרים הרבים שהתחילו ב-2012, בעקבות השקת תכנית התקשוב הלאומית "התאמת מערכת החינוך למאה ה-21" והבאת הנושא לחזית הדיון האקדמי-חינוכי בישראל. ממצא זה מדגים את הפוטנציאל של הגישה מונעת-נתונים בזיהוי מגמות שינוי בתחום טכנולוגיות הלמידה.



**תרשים 8. חמש המילים הבולטות ביותר ב-2013**

**מסקנות והשלכות**

במחקר זה נעשה שימוש ראשוני בגישה מחקרית חדשנית, מונעת-נתונים, לזיהוי מאפיינים ומגמות שינוי בשיח המחקרי בתחום טכנולוגיות הלמידה, כפי שהוא משתקף במאמרים מחקריים שהוצגו בכנסי צייס לחקר חדשנות וטכנולוגיות למידה לאורך פרק זמן של 9 שנים.

הממצאים ממחישים את הפוטנציאל הגלום בגישה מונעת-נתונים לזיהוי מונחים נפוצים בתחום טכנולוגיות הלמידה (למידה שיתופית, מורים, תלמידים), וכן תופעות בולטות בשנה ספציפית (פייסבוק, תכנית התקשוב). כמו כן, נמצא שחקירה כמותית של מילים וצירופי מילים עשויה להוביל לזיהוי תופעות מפתיעות, כגון הצירוף השכיח ביותר "פנים-אל-פנים", המשקף תמונת מחקר שלאן דווקא מתמקד בטכנולוגיה, אלא עוסק בהיבטים הפדגוגיים של שילוב טכנולוגיות בהוראה. על

סיבת האזכורים האלה לא ניתן ללמוד בסוג מחקר כזה, והאינטרפרטציה של הממצאים מחייבת שימוש בשיטות מחקר איכותניות, או כלי מחקר כמותניים נוספים.

ההנחה הבסיסית במחקר הייתה כי השיח האקדמי בתחום טכנולוגיות הלמידה הנו פועל יוצא של הטכנולוגיות הזמינות להוראה ולמידה במאה ה-21, ולפיכך הוא נתון לתמורות מהירות. ממצאי המחקר מדגימים את התרומה הפרטנציאלית של הגישה מונעת-הנתונים להבנה רטרופקטיבית של התפתחות התחום והשינויים שעוברים עליו. באמצעות גישה זו הצגנו במאמר מבט-על המהווה בסיס לבחינה עתידית של השינויים לאורך ציר הזמן, בשימוש במונחים ובמילות מפתח בתחום, ומנתונים אלה ניתן להשליך על תחומי העניין המשמעותיים של המחקר בתחום. גישה זאת עולה בקנה אחד עם מגמת השימוש הגובר במחקר שיח מונע-נתונים בעידן ה-Big Data (Radinsky et al., 2011; Radinsky & Horvitz, 2013; Michel et al., 2011).

בנוסף, לנתונים שהצגנו יש ערך כנתוני תשתית למחקרים שיתמקדו בהיבטים ספציפיים, כגון "רשת חברתית" והטכנולוגיות הקשורות אליה. שאלות מחקר ממוקדות כאלה ואחרות יכולות להיתרם מהנתונים הכמותיים שהופקו במחקר הנוכחי. לצורך כך הנגשנו את בסיס הנתונים של מאמרי ציירי מהנתונים 2006-2014 לקהילת חוקרי השיח באמצעות פלטפורמת (CQP) Corpus Query Processor (Hardie, 2012), המאפשרת חיפוש לשוני ובלשני יעיל. חוקרים המתעניינים בקורפוס זה יכולים למצוא אותו בכתובת: <http://yeda.cs.technion.ac.il/HebrewCqpWeb/learningman>

לבסוף, נשאלת השאלה עד כמה חידושי המילים, ובמידה מסוימת חידושי הנושאים הנבדקים, משפיעים על השיח ועל תוכן המחקרים בשנים שלאחר מכן. את זאת לא הצלחנו לבדוק בקורפוס הנוכחי, המתפרסם על פני תשע שנים בלבד. כמו כן, יש לזכור כי שכיחות השימוש במילים מושפעת מסגנון הכתיבה, טרמינולוגיה חדשה בתחום או מילים השייכות לזירגון מסוים.

## תודות

המאמר מציג חלק ממחקר שנערך במימון של רשות המחקר באוניברסיטה הפתוחה (מענק מספר 504292). אנו מודים לגבי דפנה עמית על הסיוע בעיבוד הנתונים, לפרופ' אלון איתי על הסיוע בהעלאת בסיס הנתונים לאתר מיל"ה (מרכז ידע לתקשוב בשפה העברית), ולגבי תהילה כרמי על ההצעה לתרגום המונח *culturomics* למונח העברי "תרבוטרקיה".

## מקורות

- לבנת, ז' (2010). *הרטוריקה של המאמר המדעי: הלשון וקהיליית השיח*. הוצאת אוניברסיטת בר-אילן.
- פרויקט השפה העברית, המרכז הארצי לבחינות ולהערכה (<http://hlp.nite.org.il/>)
- שפרן, י', בן-סימון, ע' (2014). פיתוח מדדים לאיתור חיבורים משורבטים (חיבורי גיבריש) ותיקוף המדדים מול חיבורי "פסיכורשת". דוח מרכז 401. ISBN: 978-965-502-182-0. המרכז הארצי לבחינות ולהערכה. אוחר ספטמבר 2015: <https://nite.org.il/files/reports/401.pdf>
- Anthony, L. (2011). Antconc version 3.2.4w. Center for English Language Education, Waseda University. Retrieved from: <http://www.antlab.sci.waseda.ac.jp/software.html>
- Bennet, R. E., & Ben-Simon, A. (2005). Toward theoretically meaningful automated essay scoring. Report to the National Institute for Testing and Evaluation (NITE). Retrieved from: <https://nite.org.il/files/reports/e329.pdf>
- Ben-Simon, A., & Cohen, Y. (2011). The Hebrew Language Project: Automated essay scoring & readability analysis. *The 37th International Association for Educational Assessment (IAEA) Annual Conference*, Manila. Retrieved from: [http://www.iaea.info/documents/paper\\_4e1237ae.pdf](http://www.iaea.info/documents/paper_4e1237ae.pdf)
- Ben-Simon, A., & Safran, Y. (2012). The factorial structure of written Hebrew and its application to AES. Paper presented at *The annual meeting of the National Council of Measurement in Education (NCME)*. April 12-16 2012, Vancouver, Canada. Retrieved from: <https://www.nite.org.il/files/reports/e382.pdf>
- Bohannon, J. (2011). Google Books, Wikipedia, and the future of Culturomics. *Science* 331, 135. 14 January 2011. Retrieved from: <http://www.terceracultura.net/tc/wp-content/uploads/2011/01/culturomics.pdf>

- Cuban, L. (1999). The technology puzzle. *Education Week*, 18(43), 68-69.
- Google Books Ngram Viewer. <http://books.google.com/ngrams/info>
- Hardie, A. (2012). CQPweb—combining power, flexibility and usability in a corpus analysis tool. *International journal of corpus linguistics*, 17(3), 380-409.
- Johnstone, B. 2008. *Discourse analysis*, 2nd ed. Malden, MA: Blackwell Publishing.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176-182.
- Radinsky, K., Agichtein, E., Gabrilovich, E., & Markovitch, S. (2011). A word at a time: Computing word relatedness using temporal semantic analysis. In *Proceedings of the 20th international conference on World wide web* (pp. 337-346). ACM. Retrieved from: <http://www.gabrilovich.com/publications/papers/Radinsky2011WTS.pdf>
- Radinsky, K., & Horvitz, E. (2013). Mining the web to predict future events. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 255-264). ACM.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519-549.
- Rayson, P. (2003). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison* (Doctoral dissertation, Lancaster University).
- Rayson, P. & Garside, R. (2000). *Comparing corpora using frequency profiling*. In *proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*. 1-8 October 2000, Hong Kong, pp. 1-6.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Oxford: Routledge.
- Sinclair, J. (1991). *Corpus, concordance, collocation* (Vol. 1). Oxford: Oxford University Press.
- Soper, D. S., & Turel, O. (2012a). Who are we? Mining institutional identities using *n*-grams. *System Science (HICSS), 2012 45th Hawaii International Conference on*, 1107-1116. IEEE.
- Soper, D. S., & Turel, O. (2012b). An *n*-gram analysis of Communications 2000-2010. *Communications of the ACM*, 55(5), 81-87.
- Soper, D. S., Turel, O., & Geri, N. (2014). The intellectual core of the IS Field: A systematic exploration of theories in our top journals. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 4629-4638). IEEE.
- Webber, B. (2001). Computational perspectives on discourse and dialogue. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.) (pp. 798-816). *The handbook of discourse analysis*. Oxford: Blackwell Publishers Ltd.

## נספח א: בסיס הנתונים

עשת-אלקלעי, יי, כספי, א' ויאיר, יי (עורכים). (2006). האדם הלומד בעידן הטכנולוגי: ספר כנס צייס הראשון לחקר טכנולוגיות למידה 2006, רעננה: האוניברסיטה הפתוחה, 396 עמודים (חלקית באנגלית).

<http://www.openu.ac.il/innovation/chais2006/>

עשת-אלקלעי, יי, כספי, א' ויאיר, יי (עורכים). (2007). האדם הלומד בעידן הטכנולוגי: ספר כנס צייס השני למחקרי טכנולוגיות למידה 2007, רעננה: האוניברסיטה הפתוחה, 396 עמודים (חלקית באנגלית).

[http://telem.openu.ac.il/hp\\_files/chais/07/site.html](http://telem.openu.ac.il/hp_files/chais/07/site.html)

עשת-אלקלעי, יי, כספי, א' וגרי, נ' (עורכים). (2008). ספר כנס צייס השנתי השלישי למחקרי טכנולוגיות למידה 2008, רעננה: האוניברסיטה הפתוחה, 396 עמודים (חלקית באנגלית).

[http://telem.openu.ac.il/hp\\_files/chais/08/articles.html](http://telem.openu.ac.il/hp_files/chais/08/articles.html)

עשת-אלקלעי, יי, כספי, א', עדן, ס', גרי, נ' ויאיר, יי (עורכים). (2009). **האדם הלומד בעידן הטכנולוגי: ספר כנס צייס הרביעי למחקרי טכנולוגיות למידה 2009**, רעננה: האוניברסיטה הפתוחה, 421 עמודים (חלקית באנגלית).

[http://telem.openu.ac.il/hp\\_files/chais/09/articles.html](http://telem.openu.ac.il/hp_files/chais/09/articles.html)

עשת-אלקלעי, יי, כספי, א', עדן, ס', גרי, נ' ויאיר, יי (עורכים). (2010). **האדם הלומד בעידן הטכנולוגי: ספר כנס צייס החמישי למחקרי טכנולוגיות למידה 2010**, רעננה: האוניברסיטה הפתוחה, 381 עמודים (חלקית באנגלית).

[http://telem.openu.ac.il/hp\\_files/chais/10/site.html](http://telem.openu.ac.il/hp_files/chais/10/site.html)

עשת-אלקלעי, יי, כספי, א', עדן, ס', גרי, נ' ויאיר, יי (עורכים). (2011). **האדם הלומד בעידן הטכנולוגי: ספר כנס צייס השישי למחקרי טכנולוגיות למידה 2011**, רעננה: האוניברסיטה הפתוחה, 393 עמודים (חלקית באנגלית).

[http://www.openu.ac.il/research\\_center/chais2011/papers.html](http://www.openu.ac.il/research_center/chais2011/papers.html)

עשת-אלקלעי, יי, כספי, א', עדן, ס', גרי, נ' ויאיר, יי, וקלמן, יי (עורכים). (2012). **ספר כנס צייס השביעי לחקר חדשנות וטכנולוגיות למידה 2012**, רעננה: האוניברסיטה הפתוחה, 441 עמודים (חלקית באנגלית).

<http://www.openu.ac.il/innovation/chais2012/papers.html>

עשת-אלקלעי, יי, כספי, א', עדן, ס', גרי, נ' ויאיר, יי, וקלמן, יי (עורכים). (2013). **ספר הכנס השמיני לחקר חדשנות וטכנולוגיות למידה ע"ש צייס**, רעננה: האוניברסיטה הפתוחה, 454 עמודים (חלקית באנגלית).

<http://www.openu.ac.il/innovation/chais2013/papers.html>

עשת-אלקלעי, יי, כספי, א', גרי, נ', קלמן, יי וזילבר-ורוד, וי (עורכים). (2014). **ספר הכנס התשיעי לחקר חדשנות וטכנולוגיות למידה ע"ש צייס**, רעננה: האוניברסיטה הפתוחה, 454 עמודים (חלקית באנגלית).

<http://www.openu.ac.il/innovation/chais2014/program.html>