

מבנים של מאגרי-ידע באתרי-קורסים

רפי נחמias **ארנון הרשקוביץ** **שרון הרדוף-יפה** **רונית עזרון**
 אוניברסיטת תל-אביב אוניברסיטת תל-אביב אוניברסיטת תל-אביב אוניברסיטת תל-אביב
 nachmias@post.tau.ac.il arnonher@post.tau.ac.il sharonh2@post.tau.ac.il ronit122@post.tau.ac.il

Online Repositories' Structures in Web-supported Courses

Ronit Azran **Sharon Hardof-Jaffe** **Arnon Hershkovitz** **Rafi Nachmias**
 Tel Aviv University Tel Aviv University Tel Aviv University Tel Aviv University

Abstract

The purpose of this study is to empirically reveal structures of online hierarchical repositories of knowledge items presented to university students in Web-supported courses, using data mining techniques. To this end, four variables were defined to describe the different dimensions of the hierarchical structures, and repositories of 1,203 courses were analyzed using clustering algorithm with regards to these variables. Results suggest five types of structures: over filing structure, main folder structure, flat filing structure with small folders, hierarchical structure with a pile, and flat filing structure with a pile. The article describes these structures and their frequencies on a campus-wide scale. A discussion of these results is provided.

Keywords: Knowledge repository, Hierarchical repository, Learning management systems, Data mining.

תקציר

מטרת מחקר זה לחשוף מבנים שונים של מאגרי ידע המוצגים לסטודנטים באתרים מלווי-קורסים, באמצעות שיטות של כריית נתונים. במחקר נבדקו 1,203 מאגרי ידע ולכל אחד מהם חושבו ארבעה משתנים המתארים את המבנה שלו. באמצעות ניתוח אשכולות נמצאו חמשת המבנים הבאים: מבנה של תיוק רב, מבנה של תיקייה אחת מרכזית, מבנה שטוח בעל תיקיות קטנות, מבנה היררכי בעל ערימה ומבנה של תיוק שטוח בעל ערימה. המאמר מתאר את המבנים ואת התפלגות הופעתם באוכלוסייה גדולה, ודן בממצאים.

מילות מפתח: מאגר ידע, מבנים היררכיים, מערכות לניהול למידה, כריית נתונים

מבוא

טכנולוגיית המחשוב הביאה למהפכה באקדמיה, בעיקר בעשור האחרון. מוסדות אקדמיים רבים בארץ ובעולם מאפשרים הקמה פשוטה יחסית של אתרים מלוויים לקורסים, באמצעות מערכות לניהול למידה המכונות LMS (Learning management systems), כגון HighLearn, BlackBoard ומוודל (Moodle). מערכות אלו מציעות מגוון כלים, באמצעותם מתאפשרים, בין היתר, הצגת תכנים, קיום תקשורת וניהול הלמידה. מחקרים קודמים הראו כי מרבית המרצים משתמשים באתרים אלו בעיקר להצגת תכנים ולבניית מאגר ידע היררכי עבור תלמידיהם.

מחקרים קודמים בחנו את השילוב של אתרי קורסים במוסדות להשכלה גבוהה, בדקו את יעילותם, היקפם, יתרונותיהם, ועמדות המרצים והסטודנטים אליהם. כיום, לאחר כעשור של פעילות אקדמית ברשת בישראל, ולאור תפוצתן הרבה של מערכות LMS, קיימת חשיבות לבדיקת מגוון

המבנים של מאגרי הידע בקורסים השונים העשויים להשפיע על שימושי הצריכה של הסטודנטים, ומאפשרים למפתחים להתאים את מרחבי ניהול הלמידה לסגנונות המרצים. באמצעות שיטות מחקר העושות שימוש באלגוריתמים של כריית נתונים, ניתן לבחון מבנים אלו באופן אמפירי ובהיקף רחב, באופן שלא היה אפשרי קודם לכן.

כריית נתונים (Data Mining) הינה שם כולל לאוסף של אלגוריתמים, כלים ושיטות לאיסוף וניתוח מידע אודות בסיסי נתונים גדולים, במטרה למצוא בהם דפוסים מעניינים. בשנים האחרונות, נעשה שימוש בכריית נתונים במחקר החינוכי, בין היתר במטרה לזהות דפוסי התנהגות של תלמידים במערכות למידה מבוססות-מחשב או במטרה למדוד עלות-תועלת של הוראה אקדמית משולבת-אינטרנט. מטרת מחקר זה היא להשתמש בכריית נתונים במטרה לחשוף אמפירית את המבנים ההיררכיים של מאגרי הידע המקוונים אשר מוצגים לסטודנטים בקורסים משלבי-אינטרנט.

רקע

שילובה של רשת האינטרנט בהוראה ובלמידה נראה היום טבעי ומתבקש מאליו. מאז הצגתה של הרשת הכלל-עולמית (World Wide Web) בשנת 1993, חל שינוי מהותי בתחום הלמידה, והטכנולוגיה החדשה השתלבה במהירות במהלך חיי היומיום של הלומד. טכנולוגיית המחשוב הביאה למהפכה גם באקדמיה, ובעשור האחרון החלה מגמה של הוראה המשלבת אינטרנט בלמידה באוניברסיטאות, המאפשרת לסטודנטים ללמוד מכל מקום, בכל עת ובכל נושא (Mioduser & Nachmias, 2002). בעקבות זאת, פותחו תוכנות רבות לתפעול מערכות קורסים וירטואליים, כגון: Blackboard, WebCT, Moodle, המספקות כלים להוראה ולניהול הקורס ומאפשרות למרצים שאינם בעלי רקע תכנותי לבנות אתרים עבור הקורסים שלהם (Romero, Ventura, & Garci, 2008). מערכות אלו הביאו לייעול הלמידה וההוראה בקרב המרצים באמצעות מגוון אפשרויות חדשות, כגון סיוע בפעולות אדמיניסטרטיביות (למשל, הודעות לסטודנטים, ניהול ציונים), אפשרות תקשורת (למשל, באמצעות קבוצות דיון או צ'טים), ניהול הלמידה (למשל, קיום מבחנים מקוונים) והצגת מידע באמצעות הוספת פריטי מידע ובניית מאגר ידע היררכי (Bonk, 2001). מחקרים קודמים בדקו את אופן השימוש במערכות אלו על ידי המרצים, ומצאו כי הצגת תכנים לימודיים ופרסום מידע הם שני היישומים הבולטים (Segev & Nachmias, 2003). מרבית המרצים רואים באתר הקורס מקום להצגת חומרים מגוונים, כאלו הקשורים ישירות לחומר ההוראה או מהווים חומר העשרה, והם עושים זאת באמצעות העלאת תכנים שונים, קבצי טקסט, תמונות, וידאו, סימולציות ועוד, אל מאגר הידע (נחמיאס, רם ושמלא, 2002). מערכות אלו מאפשרות למרצה לבנות מאגר פריטי ידע היררכי של תכנים, בו ניתן ליצור תיקיות, לשיים אותן ולארגן בהן קבצים, וכתוצאה מכך נוצרים מבנים רבים, שונים ומגוונים.

מרבית המחקר אודות מאגרי הייררכיים נעשה במערכות לניהול מידע אישי (Personal Information Management), המבוססות על יצירת תיקיות (וירטואליות) היררכיות ושמירת פריטים בתוכן, כמו שנעשה, לדוגמה, בתיקיית "המסמכים שלי", בתוכנת הדואר האלקטרוני וב"מועדפים" בדפדפן. במחקרים אלו זוהו שתי אסטרטגיות מרכזיות לארגון מרחב מידע אישי: (1) יצירת ערימה של פריטי מידע רבים בתיקייה הראשית; (2) יצירת תיקיות על-פי נושאים, אשר בתוכן נשמרים פריטי המידע (Malone, 1983). במחקרים נוספים נמצא כי מרבית המשתמשים עושים שימוש בשילוב של שתי אסטרטגיות אלו, ובכך יוצרים מבנים היררכיים שונים (Boardman & Sasse, 2004; Hardof-Jaffe et al., in press). בחינת מבנים של מאגרי ידע היררכיים המוצגים לסטודנטים באתרים מלווי-קורסים, תתרום רבות להבנת השימוש הנעשה באתרים אלו הלכה למעשה על ידי המרצים, ובהמשך – לאופן בו נעזרים הסטודנטים במבנים השונים.

בשנים האחרונות, החל להיעשות שימוש נרחב בכריית נתונים במחקר החינוכי, במטרה לבחון את מגוון התנהגויות הלמידה וההוראה כפי שמתבטאות בפועל באוכלוסיות גדולות (Baker & Yacef, 2009). מחקרים קודמים בחנו פְּנִים שונים של תהליך הלמידה וההוראה, כגון דפוסי למידה במערכות מקוונות וחישובי עלות-תועלת של מרכיבים מתוקשבים באתרי-קורסים (Cohen & Nachmias, 2009; Hershkovitz & Nachmias, 2009). כריית נתונים מאפשרת ניתוח רב-

מימדי באוכלוסיות גדולות, ויותר מכך: גילויים של דפוסים וקשרים חבויים ובלתי צפויים. מטרת מחקר זה היא לעשות שימוש בכריית נתונים במטרה לאפיין את המבנים השונים של מאגרי ידע היררכיים המוצגים לסטודנטים באתרי-קורסים במערכות לניהול למידה. במחקר זה נעשה שימוש בניתוח אשכולות (Cluster Analysis), תהליך בו מנתחים אוסף גדול של אובייקטים על מנת לחלקם לקבוצות ("אשכולות"), כך שהאובייקטים בתוך כל אשכול יהיו דומים זה לזה ושונים מן האובייקטים באשכולות אחרים (Han & Kamber, 2006).

מתודולוגיה

שאלת המחקר העיקרית היא: מהם סוגי המבנים השונים של מאגרי ידע המוצגים לסטודנטים באתרים מלווי-קורסים? בחינת שאלה זו תעשה באמצעות בדיקה אמפירית בקנה מידה רחב. המחקר כלל את מאגרי הידע של כל האתרים המלוויים-קורסים במערכת HighLearn (במסגרת המיזם Virtual TAU באוניברסיטת תל אביב), שהיו קיימים בסמסטר א' שנה"ל תשס"ט. מערכת HighLearn הינה מערכת לניהול למידה המאפשרת שילוב אתרי אינטרנט בקורסים אקדמיים בקנה מידה רחב ברחבי הקמפוס (נחמיאס, רם, ושמלא, 2002). במחקר נכללו אתרים שלהם לפחות 15 פריטים במאגר ידע אחד (N=1,203) מסוג פריטי תוכן בלבד (טקסט, מולטימדיה). בסך הכל, נכללו במחקר 69,092 פריטים ו-16,897 תיקיות.

לאחר איסוף הנתונים, ניקויים ועיבודם הראשוני, חושבו בעבור כל מאגר ידע 4 משתנים המתארים את מבנהו:

- שיעור הקבצים בתיקייה הגדולה ביותר מסך הקבצים במאגר הידע;
- עומק היררכי של מאגר הידע;
- גודל ממוצע של תיקייה (מספר קבצים);
- היחס בין מספר התיקיות ברמה הראשונה לעומק ההיררכי.

כתוצאה מכך, נתקבלה טבלה שבה 1,203 שורות, המתייחסות כל אחת למאגר ידע של אחד הקורסים, ו-4 עמודות, המכילות את ערכי המשתנים המתארים את המבנה שלו. על טבלה זו, הופעל אלגוריתם לניתוח אשכולות (Cluster Analysis), לזיהוי דמיון ושונות בין האובייקטים הנחקרים (מאגרי הידע). לאחר מספר איטרציות, נמצא כי מספר האשכולות הממדל בצורה הטובה ביותר את סוגי המבנים השונים הוא $k=5$. חישוב המשתנים התבצע באמצעות ניתוח האשכולות בתכנת SPSS.

ממצאים

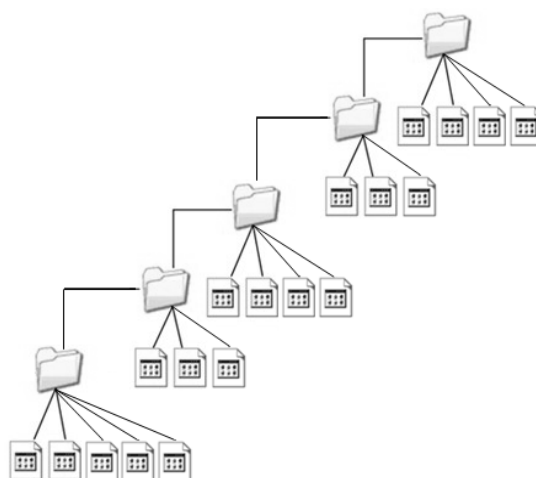
ניתוח סטטיסטי של כלל מאגרי הידע (N=1,203), העלה כי מספר הקבצים הממוצע במאגר ידע הוא 57.4 (סטטיית תקן 78.3). ממוצע מספר התיקיות במאגרי הידע עמד על 14 (סטטיית תקן 19.2), כאשר ברמה הראשונה מספר התיקיות הממוצע היה 7.9 (סטטיית תקן 6.2). עומק ההיררכיה הממוצע של מאגרי הידע הוא 1.60 (סטטיית תקן 0.89), וממוצע מס' הקבצים בתיקייה במאגר ידע הוא 5.87 בממוצע (סטטיית תקן 5.13).

כתוצאה מניתוח האשכולות, שויך כל מאגר ידע לאחת מחמש קבוצות. בטבלה 1 מופיעים, לגבי כל אחד מהמשתנים על פיהם בוצע האשכול, נתוני ממוצע וסטטיית תקן בכל אחד מן האשכולות. ערכים אלו משקפים את ההבדלים בין האשכולות שנמצאו, ולהלן תיאור חמשת המבנים השונים המיוצגים באשכולות, בצירוף איור סכמטי המתאר מבנה אופייני של מאגר ידע באשכול.

אשכול מס' 1 הוא הקטן ביותר, ומכיל 39 מאגרי ידע. **העומק ההיררכי** הממוצע באשכול זה הוא הגדול ביותר מבין ערכי הממוצע באשכולות (5.03), ובנוסף **שיעור הקבצים בתיקייה הגדולה ביותר מסך הקבצים** מקבל את הממוצע הנמוך ביותר (0.10). **ממוצע הקבצים בתיקייה** הוא 3.82. מכאן עולה כי הקבצים מחולקים לתיקיות רבות וקטנות, ונעשה שימוש במבנה ההיררכי. לפיכך, זהו **מבנה של תיקייה רב (איור 1)**.

טבלה 1. ערכי ממוצע (סטיית תקן) של ארבעת המשתנים באשכולות השונים

גודל ממוצע של תיקייה (מספר קבצים)	עומק היררכי של מאגר הידע	היחס בין מספר התיקיות ברמה הראשונה לעומק ההיררכי	שיעור הקבצים בתיקייה הגדולה ביותר מסך הקבצים במאגר הידע	גודל האשכול	מס' אשכול
3.82 (1.60)	5.03 (0.93)	2.34 (1.73)	0.10 (0.06)	39	1
17.34 (10.84)	1.01 (0.11)	1.90 (0.81)	0.92 (0.10)	82	2
2.90 (2.04)	1.09 (0.29)	15.88 (5.95)	0.16 (0.09)	180	3
4.09 (2.52)	2.18 (0.38)	3.52 (2.05)	0.24 (0.11)	444	4
6.87 (2.98)	1.05 (0.21)	5.08 (2.16)	0.49 (0.17)	458	5
5.87 (5.13)	1.60 (0.89)	5.82 (5.24)	0.36 (0.24)	1,203	סה"כ



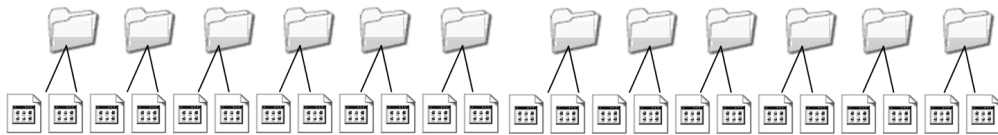
איור 1. מבנה אופייני של מאגר ידע באשכול מס' 1: מבנה של תיוק רב

באשכול מס' 2, ארבעת הממוצעים (של המשתנים המופיעים בטבלה) הם ערכי קיצון ביחס לאשכולות האחרים: שיעור הקבצים בתיקייה הגדולה ביותר מסך הקבצים במאגר (0.92) וגודל ממוצע של תיקייה (17.34) הם מירביים; היחס בין מספר התיקיות ברמה הראשונה לעומק ההיררכי (1.90) והעומק ההיררכי של מאגר הידע (1.01) הם מזעריים. מכאן שמרבית הקבצים נערמו בתיקייה אחת, ובנוסף לא נעשה כמעט שימוש במבנה ההיררכי. על כן, זהו מבנה של תיקייה אחת מרכזית (איור 2).



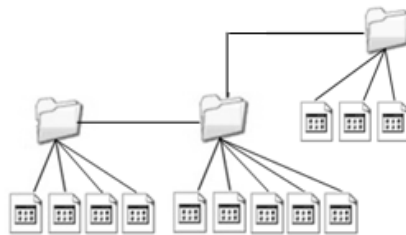
איור 2. מבנה אופייני של מאגר ידע באשכול מס' 2: מבנה של תיקייה אחת מרכזית

אשכול מס' 3 מאופיין בממוצע הגבוה ביותר של היחס בין מספר התיקיות ברמה הראשונה לעומק ההיררכי (15.88) ובגודל ממוצע של תיקייה (2.90) המקבל את הערך המזערי ביחס לאשכולות האחרים. כמו כן, העומק ההיררכי נמוך בממוצע (1.09) ושיעור הקבצים בתיקייה הגדולה ביותר מסך הקבצים במאגר (0.16) נמוך אף הוא. לפיכך, המבנה האופייני של מאגרי ידע באשכול זה הינו מבנה שטוח בעל תיקיות קטנות (איור 3), כלומר תיקיות רבות וקטנות הממוקמות בעומק היררכי נמוך.



איור 3. מבנה אופייני של מאגר ידע באשכול מס' 3: מבנה שטוח בעל תיקיות קטנות

אשכול מס' 4 מאופיין בגודל ממוצע של תיקייה שאינו גדול (4.09) ובשימוש בכ-2 רמות היררכיה, בממוצע. שיעור הקבצים בתיקייה הגדולה ביותר בו עומד על 24% בממוצע, וכך יוצא שבאשכול זה נמצאים מאגרי ידע המתאפיינים במבנה היררכי בעל ערימה (איור 4). אשכול זה הוא השני בגודלו, עם 444 מאגרי ידע.



איור 4. מבנה אופייני של מאגר ידע באשכול מס' 4: מבנה היררכי בעל ערימה

באשכול מס' 5 – שהינו הגדול ביותר, עם 458 מאגרי ידע – שיעור הקבצים בתיקייה הגדולה ביותר מסך הקבצים במאגר הידע עומד על כחצי. העומק ההיררכי נמוך בממוצע (1.05) והתיקות מתאפיינות בגודל בינוני בממוצע (6.87). לכן, מאגרי הידע באשכול זה מתאפיינים בשילוב בין תיוק לעירום ושימוש נמוך במבנה ההיררכי. כלומר, זהו מבנה של תיוק שטוח עם ערימה (איור 5).



איור 5. מבנה אופייני של מאגר ידע באשכול מס' 5: מבנה של תיוק שטוח עם ערימה

בדיקות ההתפלגות של חמשת המבנים שנמצאו, העלתה כי שני המבנים הנפוצים הם: מבנה של תיוק שטוח עם ערימה (38%) ומבנה היררכי בעל ערימה (37%). מבנים אלו משלבים את שתי האסטרטגיות הבסיסיות של ארגון מידע: תיוק ויצירת ערימות. ממצאים דומים עולים במחקרים שבדקו ניהול מידע אישי ומצאו כי מרבית המשתמשים משלבים בין שתי אסטרטגיות אלו (Whittaker & Hirschberg, 2001; Boardman & Sasse, 2004). שאר מאגרי הידע התחלקו בין שלושת המבנים האחרים, 15% שייכים למבנה שטוח בעל תיקיות קטנות, 7% מאופיינים כמבנה של תיקייה אחת מרכזית ורק 3% מהקורסים מוגדרים כמבנה של תיוק רב.

דיון

כיום, אתרים מלווי-קורסים הם דבר שבשגרה באקדמיה; כך, למשל, באוניברסיטת תל אביב כ-75% מן הקורסים הם בעלי אתר-אינטרנט מלווה (Nachmias & Ram, 2009). השימוש המרכזי אשר עושים אתרים אלו הוא בהצגת תכנים (Nachmias & Segev, 2003; Shemla & Nachmias, 2007), ולכן צורות הארגון השונות של התכנים המקוונים בקורסים אלו מהוות את פניה של ההוראה האקדמית המתקדמת. מחקר זה חושף את המבנים השונים של מאגרי הידע המוצגים לסטודנטים באתרים מלווי-קורסים ואת התפלגותם בפועל, ומאפשר להבין יותר אודות תוצריה של הפעילות המקוונת. הבנה זו עשויה להיות לעזר רב הן לקובעי-מדיניות באוניברסיטאות, בכל הנוגע

להטמעת מערכות המותאמות יותר לאוכלוסיית המשתמשים, והן למפתחי מערכות להוראה מתוקשבת, בבניית מרחבי למידה המתאימים לסגנונות הארגון של המרצים ולהעדפותיהם של התלמידים. בכך, מעשיר מחקר זה את הידע אודות השימושים במערכות לניהול למידה, כפי שהצטבר בשנים האחרונות במחקרים רבים אשר בחנו פנים שונים של ההוראה והלמידה במערכות אלו ובמתודולוגיות שונות (לדוגמא, Cohen & Nachmias, 2009; Shemla & Nachmias, 2007).

במחקר זה נעשה שימוש במתודולוגיה של כריית נתונים, אשר אפשרה לחשוף את המבנים המורכבים והמגוונים של מאגרי הידע. בשנים האחרונות, נעשה שימוש הולך וגדל במתודולוגיה זו במחקרים חינוכיים, במטרה לזהות דפוסים ותופעות הקשורים ללמידה מתוקשבת. השימוש שנעשה כאן במתודולוגיה זו מדגים כיצד איסוף ועיבוד נתונים באמצעות כריית נתונים מאפשר ניתוח רב-מימדי באוכלוסייה גדולה וחושף דפוסים חבויים בעלי משמעות.

היווצרותם של מבנים שונים של מאגרי ידע באתרי-קורסים עשויה להיות קשורה למאפיינים שונים של הקורסים, כגון: גודל מאגר הידע, שיוך פקולטטי ומספר הסטודנטים בקורס; בניית עתידי, נבחן קשרים אלו. בנוסף, חשוב לבחון בהמשך האם קיימים קשרים בין מבנה מאגר הידע ואופן השימוש בפריטי התוכן שבו על ידי הסטודנטים, כפי שבא לידי ביטוי במשתני צריכה שונים. המושך מחקר בכיוונים אלו ירחיב ויעמיק את הבנתנו את השימוש במערכות למידה מתוקשבות הנהוגות כיום באקדמיה, ובכך עשוי לתרום להגברת יעילותן של מערכות אלו.

מקורות

- נחמיאס, ר', רם, י', ושמלא, ע' (2002). למידה אקדמית ברשת באוניברסיטת תל-אביב – היבטים פדגוגיים ומסקנות ראשוניות. **על הגובה**, גיליון 1, עמ' 38-41.
- Baker, R.S.J.D. and Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Boardman, R. and Sasse, M. A. (2004). "Stuff goes into the computer and doesn't come out": A cross-tool study of personal information management. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, 24-29 April 2004*, 583-590.
- Bonk, C.J. (2001). *Online teaching in an online world*. Bloomington, IN: CourseShare.com. Retrieved October 28, 2009, from http://php.indiana.edu/~cjbbonk/faculty_survey_report.pdf.
- Cohen, A. & Nachmias, R. (2009). Quantitative Cost Effectiveness Model for Web-Supported Academic Instruction. In Bornath, U., Szucs, A., Tait, A., & Vidal, M. (Eds.), *Distance and E-learning in Transition – Learning innovation, technology and social challenges*. ISTE and Wiley & Sons, Inc, 151-160.
- Han, J. & Kamber, C. (2006). *Data mining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann Publishers.
- Hardof-Jaffe, S., Hershkovitz, A., Abu-Kishk, H., Bergman, O., & Nachmias, R. (in press). Students' organization strategies of personal information space. To appear in *Journal of Digital Information*.
- Hershkovitz, A. & Nachmias, R. (2009) Learning about online learning processes and students' motivation through Web usage mining, Interdisciplinary. *Journal of E-Learning and Learning Objects*, 5(2009), 197-214.
- Malone, T. W. (1983). How do People Organize their Desk? Implications for the Design of Office Information Systems. *ACM Trans. on Office Information Systems*, 1(1), 99-112.
- Mioduser, D. & Nachmias, R. (2002). WWW in education. In H. H. Adelsberger, B. Collis & J. M. Pawlowski (Eds.), *Handbook on Information Technologies for Education and Training* (pp. 22-63). Berlin: Springer-Verlag.
- Nachmias, R. & Ram, J., (2009). Research Insights from a Decade of Campus-Wide Implementation of Web-Supported Academic Instruction at Tel Aviv University. *International Review of Research in Open and Distance Learning*, 10(2), 1492-3831.

-
- Nachmias, R. & Segev, L., (2003). Students' Use of Content in Web-Supported Academic Courses. *The Internet and Higher Education*, 6(2), 145-157.
- Romero, C. Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384.
- Shemla, A. & Nachmias, R. (2007). Current State of Web-Supported Courses at Tel-Aviv University. *International Journal of E-Learning*, 6(2), 235-246 .
- Whittaker, S. & Hirschberg, J. (2001). The character, value, and management of personal paper archives. *Transactions of Computer Human Interaction*, 8(2), 150-170.