

אוניברסיטת חיפה

החוג לניהול מידע וידע

מספר הקורס : 218.4030

שם הקורס: אחזור מידע (Information Retrieval)

שם המרצה: פרופ' צבי קופליק

שעות קבלה: יום ג', בין השעות 9 ל 12, חדר TBD

דואר אלקטרוני: [Email: tsvikak@is.haifa.ac.il](mailto:tsvikak@is.haifa.ac.il)

#### מטרת הקורס

הקורס יציג עקרונות, גישות וטכניקות באחזור מידע. מחקר באחזור מידע (טקסטואלי) החל כבר בשנות ה 40 של המאה ה 20, כאשר המטרה העומדת מאחורי אחזור הייתה לשמר מבחר נרחב של טקסט ( ספרים, מסמכים, ומאמרים) ולאפשר חיפוש חופשי של אוספים אלה. האתגרים אשר עומדים ועמדו בפני החוקרים היו מקום אחסון, יעילות, דיוק ועוד. עם התפתחות תקשורת המחשבים והגישה המקוונת למאגרי מידע ובהמשך, עם צמיחתה והתפתחותה של רשת האינטרנט התעורר מחדש העניין באחזור מידע. כמות המידע הקיים ברשת דורש שיטות חיפוש יעילות אחרת כל חיפוש יהפוך לחסר תועלת. הקורס מציג טכניקות כלליות לניתוח וייצוג טקסט, טיפול בשאליות, הערכת מערכות אחזור מידע ויישומים שונים של טכניקות אלו.

#### פירוט נושאים ולוח זמנים

1. שבוע ראשון, הרצאת מבוא: הצגת מבנה הקורס ואופן הלימוד וכן הרצאת מבוא באחזור מידע, אחזור מידע על פי המודל הבוליאני ופעולות בסיסיות על טקסט.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.
2. שבוע שני: נושאי המשך בפעולות על טקסט, בנייה ודחיסת אינדקס.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.
3. שבוע שלישי: שקלול מונחים/מלים והמודל הווקטורי.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.

4. שבוע רביעי: הערכת תוצאות אחזור מידע.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.
5. שבוע חמישי: משוב משתמש והרחבת שאילתות.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.
6. שבוע שישי: סיווג טקסט ויישומו באמצעות מודלים שונים.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.
7. שבוע שביעי: טכניקת אגדים/אשכולות לקיבוץ טקסט.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.
8. שבוע שמיני: חיפוש מידע באינטרנט.  
<<< מטלה שוטפת לשיעור הבא – שאלות על פרקי הלימוד.

#### ספרות

הקורס עוקב אחר ספר הלימוד: An introduction to Information retrieval, מאת C. Manning, P. Raghavan, H. Schütze הזמין בגרסת טיוטה באופן מקוון ב: <http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>  
ספר עזר נוסף בנושא רקע מתימטי לאחזור מידע הוא ספרו של Dominich Sandor :  
Mathematical Foundations of Information Retrieval  
ספרי עזר נוספים באחזור מידע:  
Modern Information Retrieval מאת Baeza-Yates & Ribeiro-Neto  
Information Retrieval, Algorithms and Heuristics, 2<sup>nd</sup> ed. מאת Ophir Frieder ו David Grossman

בנוסף חומר רב מאוד זמין באופן חפשי באינטרנט ויעשה בו שימוש במהלך הקורס.

#### מטלות הקורס והרכב הציון

מטלות הקורס יכללו תרגילים לצורך תרגול של חלק מן החומר הנלמד בכיתה, בצמוד לשעורים, מטלה מעשית ומבחן מסכם.  
משקל המבחן 100%, חובת מעבר המבחן בציון 65 ומעלה.  
משקל התרגילים השוטפים 0% (לתרגול שלכם), אולם קיימת חובת הגשה כתנאי להשתתפות במבחן.  
הגשת כל המטלות היא חובה ותנאי לגישה למבחן!  
אי הגשת תרגילים/מטלות פירושה אי עמידה בדרישות הקורס.

#### השגת מטרות הקורס



החוג לניהול מידע וידע  
The Department of Information  
and Knowledge Management



הקורס מלווה במטלות שוטפות ובתרגילים באופן שוטף. מטרת המטלות השוטפות לסייע בהטמעת החומר הנלמד באופן שוטף. מטרת התרגילים לספק התנסות מעשית לצורך המחשה של נושאי הלימוד, בצמוד ככל האפשר לנושא הלימוד הרלוונטי. מטלות הקורס פורטו במסגרת תכנית הקורס ורוכזו בסעיף הקודם.