



# קובץ תקצירים כנס חצי היובל קבוצת עניין אחזור מידע וטקסט – SIGTRTS

בית הספר להנדסה ומדעי המחשב,  
האוניברסיטה העברית בירושלים  
אוקטובר 2018

1. קול קורא לכנס
2. תוכנית הכנס
3. פתיחה : עולם אחזור המידע בחצי היובל כפי שהוא בא לידי ביטוי בקבוצת העניין אחזור מידע וטקסט – SIGTRTS, ד"ר עפר דרורי, שע"ם
4. אחזור מידע : אתמול, היום ומחר (?), פרופ' יעקב שויקה, אוני' בר אילן
5. חדשנות טכנולוגית, ד"ר גיימי שוורצקופ, STKI
6. נפלאות הבינה המלאכותית, מאור דוד-פור, מיקרוסופט
7. מאחזור מידע לכריית טקסט – טוויטר כמקור מידע עדכני להלכי רוח הציבור, פרופ' צבי קופליק, אוני' חיפה
8. סיכום

## קבוצת ענין אחזור מידע וטקסט - SIGTRS



כ"ב אייר תשע"ח  
07 מאי 2018

לכבוד  
חברי קבוצת ענין אחזור מידע וטקסט - SIGTRS  
שלום רב

### "קול קורא" לכנס חצי היובל של קבוצת העניין אחזור מידע וטקסט – SIGTRS, 10/10/18

קבוצת העניין אחזור מידע וטקסט - SIGTRS מהווה פורום לאנשי מקצוע העוסקים בתחום אחזור מידע, טקסט וטכנולוגיות קשורות. הקבוצה הוקמה בשנת 1994 ע"י החתום מטה לאור צורך שאותר לטובת העוסקים בתחום ובמגמה של שתוף מידע לא על בסיס מסחרי. מזה עשרים וחמש שנים נפגשת הקבוצה בקביעות ארבע פעמים בשנה להרצאות ולעלות רעיונות. הקבוצה מפרסמת בכל השנים עלון ששמו "מידע וטקסט" ובו מידע מלא על כל ההרצאות שהתקיימו בקבוצה כולל טקסט מלא של ההרצאות והמצגות שהועברו. אחרי 25 שנות פעילות מצאתי לנכון לקיים כנס שיהווה נקודת ציון לחצי יובל של פעילות. הכנס אמור להתקיים בבית הספר להנדסה ולמדעי המחשב של האוניברסיטה העברית בתאריך 10 באוקטובר 2018 למשך של כחצי יום. חלק מההרצאות אושרו ע"י בכירים בתחום ואני מאפשר להגיש עוד הרצאות ומאמרים למפגש זה. אודה על קבלת רעיונות להרצאות בצרוף תקציר של עד עמוד אחד אשר יידרש להיות מורחב במידה וההצעה תתקבל. נא להעביר את ההצעות והתקציר לעפר דרורי ב- [offerd@gmail.com](mailto:offerd@gmail.com) בבקשה שרינו יומנים וציינו במפורש את כוונתכם להגיע בצרוף מספר פלאפון מספר המקומות מוגבל. תודה ולהתראות בכנס.

בברכה

עפר דרורי  
יו"ר הקבוצה

# קבוצת ענין אחזור מידע וטקסט - SIGTRS



כ"ג אלול תשע"ח  
03 ספטמבר 2018

לכבוד

חברי קבוצת ענין אחזור מידע וטקסט - SIGTRS  
שלום רב

**כנס חצי היובל של קבוצת הענין אחזור מידע וטקסט – SIGTRS, 10 באוקטובר 2018**

**בית הספר להנדסה ומדעי המחשב האוניברסיטה העברית בירושלים**

**אולם B-221**

## תוכנית הכנס

הרשמה וכינוס	0900 - 0930
פתיחה: עולם אחזור המידע כפי שהוא בא לידי ביטוי בקבוצת הענין אחזור מידע וטקסט – SIGTRS, עפר דרורי, שע"ם	0930 – 1000
אחזור מידע: אתמול, היום ומחר (?), פרופ' יעקב שויקה, אוני' בר אילן	1000 – 1100
חדשנות טכנולוגית, ד"ר גימי שוורצקופ, STKI	1100 – 1200
הפסקה	1200 – 1215
נפלאות הבינה המלאכותית, מאור דוד-פור, מיקרוסופט	1215 – 1300
מאחזור מידע לכריית טקסט – טוויטר כמקור מידע עדכני להלכי רוח הציבור, פרופ' צבי קופליק, אוני' חיפה	1300 – 1345
סיכום	1345 – 1400

להרשמה יש לשלוח דוא"ל לעפר דרורי יו"ר הקבוצה – [offerd@gmail.com](mailto:offerd@gmail.com) בצרוף שם מלא, ארגון, דוא"ל ומספר פלאפון. מספר המקומות מוגבל. הכניסה לכנס מותנית בהרשמה מראש.

קבוצת הענין אחזור מידע וטקסט - SIGTRS מהווה פורום לאנשי מקצוע העוסקים בתחום אחזור מידע, טקסט וטכנולוגיות קשורות.

הקבוצה הוקמה בשנת 1994 ע"י החתום מטה לאור צורך שאותר לטובת העוסקים בתחום ובמגמה של שתוף מידע לא על בסיס מסחרי.

מזה עשרים וחמש שנים נפגשת הקבוצה בקביעות ארבע פעמים בשנה להרצאות ולעלות רעיונות. הקבוצה מפרסמת בכל השנים עלון ששמו "מידע וטקסט" ובו מידע מלא על כל ההרצאות שהתקיימו בקבוצה כולל טקסט מלא של ההרצאות והמצגות שהועברו.

כנס חצי היובל נועד לציין 25 שנות פעילות של הקבוצה כנס שיהווה נקודת ציון לחצי יובל של פעילות.

בברכה

עפר דרורי

מייסד ויו"ר קבוצת ענין אחזור מידע וטקסט SIGTRS

# עולם אחזור המידע בחצי היובל כפי שהוא בא לידי ביטוי בקבוצת העניין אחזור מידע וטקסט – SIGTRS

## עפר דרורי, שע"ם

בראשית היו מנועי חיפוש אבל קדמו להם המדריכים. בעולם ראשוני בו החלו להצטבר ברחבי הרשת מסמכים רבים והדרך לאתר אותם הייתה מוגבלת פותחו המדריכים. תוכנות שיצרו היררכיה נושאית על מרבית המידע ברשת כאשר המחפש של המידע ניווט עצמו באמצעות אתרי המדריכים למידע שרצה. לא היה חיפוש של מסמך מסוים כפי שנהוג היום אלא חיפוש נושאי. התעניינות במידע תיירותי על צפון איטליה? הדרך להגיע אליו הייתה דרך קטגוריה של "תיירות" ממנה לדלג לקטגוריה "אירופה" משם לקטגוריה "איטליה" משם ל-"צפון איטליה" ומשם לפי הצורך, "מלונות" לרשימת מקומות לינה או "אגמים" לרשימת אתרי מים וכדומה.

משמעות השיטה הייתה קטלוג ידני של כלל המידע ברשת דבר שכמובן לא היה אפשרי אבל סיפק פתרון לא רע בשנים הראשונות של האינטרנט. בנוסף לא הייתה בהכרח התאמה בין המונח/מושג "ההגיוני" בעיני המחפש לבין זה שעליו חשב מקטלג האתרים. הדבר היה נכון שבעתיים בעת שימוש באתרים בשפות שונות שבהם מיומנות הגדרת הקטגוריות בשפות שונות היא קריטית. שיטת חיפוש זו מתאימה לאנשים עם סבלנות וסקרנות טבעית רבה אבל פחות מתאימה לאנשים קצרי רוח ותזזייתיים המאפיינים את הדור שלנו.

מנועי החיפוש ובראשם גוגל יצרו ליגה חדשה. לא עוד ניווט נושאי אלא כתיבה ישירה של הנושא המעניין אותי כמו למשל "מלון בצפון איטליה ליד אגם".

בהמשך התפתחו המנועים שניסו לסייע למחפש בניחוש רצונו. פעולה זו בוצעה באמצעות ניחוש מונחי החיפוש ע"י השלמת מחרוזת החיפוש בהתבסס על חיפושים קודמים, בקטגוריזציה של התוצאות וחיבורן באשכולות נושאיים ועוד ועוד. שיטה זו אפשרה לסוחרים גם להציע אתרים דומים או מוצרים דומים שאפשרה גם הגדלת המכירות אבל לעתים גם סייעה אמתית באיתור מידע שהתקשינו להגדיר בעת החיפוש.

אנו נמצאים בעולם בו נותני השירותים מנסים לנחש את התשובה לשאלה שטרם נשאלה, לעתים זה מטרד ולעתים מסייע, הכל תלוי בדרך ובאחוז הצלחת הניחוש.

בהרצאה נדבר על מגמות בעולם החיפוש וגם נזכיר את פעילות קבוצת העניין לאחזור מידע וטקסט SIGTRS בעשרים ואחת השנים האחרונות.

## על המרצה

ד"ר עפר דרורי, מייסד אתר קבוצת עניין אחזור מידע וטקסט SIGTRS, מייסד ומנהל את אתר הגבורה המביא מידי על כלל מקבלי הצ"שים והעיטורים בכוחות הביטחון. מייסד מתודולוגיית שילוב לניהול מחזור חיים של מערכות מידע וסגן מנהל יחידת מחשב גדולה. עפר דרורי כתב עד כה תשעה עשר ספרים ויותר משלוש מאות מאמרים.



## אחזור מידע: אתמול, היום ומחר (?) פרופ' יעקב שויקה, אוני' בר אילן

בהרצאה זו נסקור מספר תופעות בולטות בהתפתחותו של התחום הזה – אחזור מידע – מאז שנות הששים המוקדמות, תוך אזכור אפיזודות אישיות שהמחבר חווה בלוותו את התחום מקרוב לאורך השנים, ותוך שימת דגש על מורכבות השפה העברית והבעיות המיוחדות שהיא מעוררת, והגישות האפשריות לפתרון.

### על המרצה

יעקב שויקה, פרופסור אמריטוס למדעי המחשב, הוא מחלוצי חקר אחזור מידע בישראל, היה ראש פרויקט השו"ת בבר-אילן בשנים 1974 – 1985 וראש המכון לאחזור מידע ובלשנות חישובית שם (הפרויקט זכה בפרס ישראל ב-2007), פיתח סדרת תשתיות למחשוב השפה העברית, ובעשור האחרון עמד בראש סדרת פרויקטים שאפתניים למחשוב במדעי היהדות, ובהם מחשוב אוסף כתבי היד של גניזת קהיר, ובניית אתר לריכוז והצגה סינופטית של כל שינויי הנוסחאות בתלמוד הבבלי לפי כל כתבי היד בעולם.



## כיצד ניתן לייצר סוגים שונים של חדשנות ומתודולוגיות לניהול מוצלח של תוצריה? דר' ג'ימי שוורצקופ STKI

המצגת עוסקת בבנייה ויישום של תרבות החדשנות הקמה והנחלה של מתודולוגיות לניהול החדשנות כגון: Design Thinking, בניית מודלים עסקיים חדשים, מעבדות חדשנות, יצירת Ecosystem המאפשר שיתוף פעולה חכם עם אנשים יצרניים ושירותים.

### על המרצה

מנכל STKI ומרצה לחדשנות במכללה האקדמית ת"א יפו.



## **נפלאות הבינה המלאכותית**

**מאור דוד-פור, מיקרוסופט**

איך אפשר להפוך את טכנולוגיית הבינה המלאכותית לנחלת הכלל?  
בינה מלאכותית מתקיימת בשורה של מוצרים ושירותים טכנולוגיים, שהופכים יותר ויותר שגורים בשימוש היום-יומי. שירותי שפות ותרגום, עוזרות קוליות, זיהוי תמונה ווידאו - כולם נסמכים על למידת מכונה ולמידה עמוקה.  
בהרצאה תוצג הטכנולוגיה שמאחורי הבינה המלאכותית, את היכולות שלה ואת המגבלות שלה, וננסה לנחש איך יראה העתיד הקצת יותר רחוק.

### **על המרצה**

מנהל מוצר Microsoft 365 במיקרוסופט ישראל, מתכנת וארכיטקט לשעבר עם תשוקה עזה לטכנולוגיה ואנשים.





**מאחזור מידע לכריית טקסט – טוויטר כמקור מידע עדכני להלכי רוח הציבור – תהליך כללי לחילוץ מידע מטוויטר והדגמתו בתחומי התחבורה וסדרות טלוויזיה פרופ' צבי קופליק, אוני' חיפה**

אם את שרשי אחזור המידע משייכים לשנות ה 40 של המאה ה 20, את המונח כריית טקסט נתין למצוא כבר בשנות ה 70. אחזור מידע ובהמשך כריית טקסט זכו לפריחה ועניין רב עם הופעת האינטרנט וביתר שאת עם התפתחות Web 2.0 וה Semantic Web, כאשר כולנו הפכנו ליוצרי תוכן. רשת האינטרנט הפכה למקור מידע חשוב כמעט בכל תחום. בהיבט זה, הרשת משקפת את עמדות הציבור במגוון נושאים ובפרט המדיה חברתית ובתוכה טוויטר – "המערכת הפצת הודעות מידיות באינטרנט". טוויטר מאופיין בהודעות קצרות – "ציוצים" המשמשים במקרים רבים לדיווח מידי – אודות אירועים ו/או הבעת דעות. במסגרת מחקרית בחנו את האפשרות להפיק תובנות המעניינות בעלי עניין בשני תחומים שונים: תחבורה וסדרות טלוויזיה. המחקר המדווח המוצג מציע מסגרת כללית לכרייה מידע מטוויטר ומודגם על הפקת תובנות מכריית טקסט הקשור לתחבורה במקרה אחד וסדרות טלוויזיה במקרה שני. מוצגים האתגרים אליהם יש להתייחס לפני יישום נרחב של מסגרת זו לצורך קצירה והפוטנציאל להפקת תובנות על ידי בעלי עניין.

**על המרצה**

**צבי קופליק** הוא פרופסור מן המניין באוניברסיטת חיפה, לשעבר ראש החוג למערכות מידע וכיום ראש שותף של תכנית הלימודים לתואר ראשון במדעי הרוח הדיגיטליים. מחקרו מתמקד בממשקי אדם מחשב חכמים ובין היתר בהפקת תובנות מתוכן טקסטואלי של מדיה חברתית, הנושא אותו יציג בהרצאה.





## דברי סיכום

מטרת כנס חצי היובל לקבוצה הייתה להוות ציון דרך לפעילות מתמשכת של עשרים וחמש שנים ברציפות.

תחילתה בהצעה של מייסד הקבוצה עפר דרורי לארגון משתמשים של בית תוכנה גדול (SPL) ולאחריה בקשה של אותו בית תוכנה להקים את הקבוצה לתועלת משתמשי הארגונים שאתם היה קשור בית התוכנה בקשרים עסקיים.

לאחר זמן קצר נוסדה הקבוצה ופעילותה עברה מחסות של ארגון המשתמשים לפעילות עצמאית של הקבוצה לתועלת העניין של אנשי עולם אחזור המידע בארץ תוך התמקדות בייחודיות של השפה העברית ולא רק.

מעט נתונים מסמכים על פעילות הקבוצה :

- \* נוסדה בשנת 1994
- \* פורום פתוח לאנשי מקצוע בתחום אחזור המידע וטכנולוגיות קרובות
- \* 4 מפגשים בשנה, 90 מפגשים עד היום
- \* עלון "מידע וטקסט" מופיע פעמיים בשנה
- \* 48 חוברות ב-25 כרכים כ-200 מאמרים פורסמו עד כה
- \* כשמונה מאות חברים רשומים
- \* אתר הכולל טקסט מלא של כל ההרצאות והמאמרים ב 25 השנים האחרונות

הכנס התקיים בבית הספר להנדסה ומדעי המחשב באוניברסיטה העברית בירושלים ואנו מודים להנהלת בית הספר על האכסניה הנאה והנוחה. זו גם הזדמנות טובה להודות לשרות עיבודים ממוחשבים על האכסניה הטובה והרציפה בשנים הרבות של פעילות הקבוצה.

אני מודה למרצים ולחברי הקבוצה המתמידים להגיע למפגשים וגם לאלו הקוראים מרחוק את החומר. להתראות במפגש הבא

עפר דרורי

מייסד קבוצת עניין אחזור מידע וטקסט – SIGTRS  
יו"ר כנס חצי יובל לפעילות הקבוצה