



כח' באייר תשע"ט
2 ביוני, 2019

"מידע וטקסט"

עלון קבוצת עניין אחזור מידע וטקסט - SIGTRTS
חוברת 1 כרך כו' – יוני 2019

חדשות מקבוצת העניין

שלום לכולם!

אנו נפגשים שוב בעלון חדש נוסף. עלון זה פותח את כרך כו' של הקבוצה. אני מאחל לכולנו המשך הנאה, עניין ופעילות ברוכה בתחום אחזור המידע.

ראוי לציין שנושאים רבים הנידונים בקבוצה נקבעים ע"י החברים עצמם בדפי המשוב בסיום המפגשים. אתם מוזמנים להמשיך ולהציע נושאים לדיון וכמובן להציע את עצמכם להרצאה.

אני מזכיר לכם את כתובת האתר של הקבוצה www.sigtrts.org. באתר חומרים רבים שנאספו במשך שנים רבות בתחום. אנא הפיצו בין חבריכם את כתובתנו ועודדו אותם להצטרף לרשימת התפוצה (הרשימה מפוקחת על ידי והתעבורה בה נועדה לעדכונים בלבד).

1. קשר

הקבוצה מקימת קשר עם חבריה באמצעות קבוצת דיוור אלקטרונית (Mailing-list) המאפשרת בצורה חופשית לכל אדם להירשם לקבוצה או למחוק עצמו ממנה. הדרכה כיצד להצטרף או לעדכן כתובת מופיעה באתר תחת התפריט "רשימת תפוצה של הקבוצה".

ספריות וגופים מרכזיים אחרים המעוניינים לקבל עותק מודפס של העלון צריכים לפנות בבקשה מיוחדת ליו"ר הקבוצה.

הרוצים להיות חברים בקבוצה צריכים להירשם לרשימת התפוצה האלקטרונית שלה באתר הקבוצה.

באתר הקבוצה מפה שנועדה להקל על ההגעה של החברים מחוץ לעיר. הדפיסו אותה לפני היציאה. אם אתם עושים שימוש בניווט לווייני (GPS) כונו אותנו לשע"ם, רחוב פועלי צדק 4, ירושלים. נסיעה טובה!

מאז הוצאת החוברת האחרונה (כרך כה' חוברת מס. 2) בדצמבר 2018 נפגשה הקבוצה פעמיים.

המפגש הראשון התקיים בינואר 2019

א. הרצאה ראשונה: מנוע לגילוי סרטים מבוסס ניתוח שפה טבעית ע"י יורם שפר הוצגה המוטיבציה לאיתור סרטים לנושא מסוים. במהלך בניית המאגר נעשו החלטות על סמך איזה מידע נכון להתבסס לצורך קטלוג החומר: ניתוח הסרט עצמו, תקציר הסרט, תגיות וכו'. ההחלטה היתה לעשות שימוש בתקצירי הסרטים 10 (ד' ראשונות). הוצג המנוע ודרך פעולתו. הוצגה הייחודיות של המנוע הן במידע איכותי ביותר של התוצרים, אלגוריתם החיפוש, דיוק בתוצאות.

ב. הרצאה שנייה: איתור ישויות מתוך אוכלוסייה רחבה ע"י אביבה יפת בהרצאה הוצגו ההתמודדות עם מקורות מידע רחבים ומגוונים כאשר המשימה המרכזית בטיוב המידע היא הצמדת כל מאות מיליוני פרטי המידע לישויות הרלוונטיות הנכונות. הוצגו הדרכים השונות לאיתור שגויים למיניהם, ביצוע תקנון ואחידות בשמות, כתובות ושאר פרטים מזהים ומתוך שקלול כל המידע ניתן להגיע בוודאות גבוהה לישויות הנכונה. הודגם השימוש בכלי.

המפגש השני התקיים אפריל 2019

א. הרצאה ראשונה: המידע הגאוגרפי באתר ממשלתי בחזית ומאחורי הקלעים ע"י אלכס קורן ניתנה סקירה על המידע הקיים בארגון ועל התהליך המתוכנן למעבר המידע לענן. הפיתוח נעשה קודם כל עבור ציוד נייד (פלאפונים וכדו') ורק לאחר מכן למחשבים שולחניים. הוצגו השרותים הקיימים היום במערכת כולל שרותים חיצוניים הניתנים באמצעות API.

ניתן להתרשם מהמערכת ותכונותיה בגישה ישירה לאתר Gov Map

ב. הרצאה שנייה: ניתוח מוכוון נתונים של שיח פוליטי (ניתוח סמנטי של טקסטים) ע"י יאיר פוגל דרור יאיר הציג את ההקשר של העבודה שלו והוא ניתוח טקסטים בהקשר פוליטי במסגרת עבודת הדוקטורט שלו. הפתרון שפותח להבנת הטקסט הוא חלוקת התהליך לשלבים. בשלב א' מתייגים את הטקסט לקטגוריות בהתחלה בצורה ידנית ע"י אנשים ולאחר מכן בצורה ממוחשבת. בשלב ב' החומר שתוייג מוכנס לתהליך ממוחשב המייצר קטגוריות. הוצגו דוגמאות לניתוח קורפוסים של טקסט עיתונאי מהתקשורת והופעלו עליהם הכלים תוך ניתוח והבנה של הטקסט ושיוכם לנושאים מוגדרים מראש.

מאמרים ועבודות

א. גישה לנתונים Operational Data Store ע"י פיני כהן

ב. Robotic Process Automation - RPA ע"י גלית פיין

ג. הבלשנית שהסבירה: אם אין פלסטין, אין ישראל ע"י עופר אדרת

קריאה מהנה.

4. חסות

נכון להיום הקבוצה ללא חסות.
נשמח לקבל הצעות לחסות מארגונים בתחום העיסוק של הקבוצה.

5. כללי

קבוצת העניין "אחזור מידע וטקסט" (SIGTRS) מהווה פורום לאנשי מקצוע העוסקים בתחום אחזור טקסט, אחזור מידע וטכנולוגיות קשורות. אנשי המקצוע הם מפתחים (מנתחי מערכות ותוכניות) ו/או משתמשים.
הקבוצה הוקמה בשנת 1994 ע"י החתום מטה ומאז פועלת ברציפות הן במפגשי הרצאות והן בהפצת מידע בין היתר באמצעות עלון הקבוצה.
הקבוצה נפגשת ארבע פעמים בשנה להרצאות ולהחלפת רעיונות.
עלון הקבוצה יוצא פעמים בשנה בקביעות משנת 1994. כל חומר העלון בטקסט מלא נמצא באתר הקבוצה.
הצעות להרצאות, רעיונות או חומר כתוב אחר ניתן לשלוח ליו"ר הקבוצה :

עפר דרורי

שע"ם

ת.ד. 10414, ירושלים 91103

טל: 02-5688439

או בדואר אלקטרוני - offerd@gmail.com (הכי טוב...)

ב ב ר כ ה

עפר דרורי
יו"ר הקבוצה

בקרו אותנו באתר הבית של הקבוצה : <http://www.sigtrs.org>