



"מידע וטקסט"

עלון קבוצת עניין אחזור מידע וטקסט - SIGTRTS

1. חדשות מקבוצת העניין / עפר דרורי.....1
- מפגשי הקבוצה:**
2. מבוא לזיהוי דיבור - עידן בן שלוש.....5
3. טכנולוגיה סמנטית להבנת שפה - שאול שקד.....9
4. גרפים קונספטואליים (CGS) דרך מעניינת להבנה סמנטית - גיל פוקס.....15
5. פרמטרים המסייעים להצלחת דיבור מוחשב - דורון מודן.....23

מאמרים:

לא נמסרו בחברת זו מאמרים לפרסום.

תוכן עניינים

6. אינדקס לכרכים א' עד כ' (כולל חוברת 2) עפ"י מחברים.....26
7. אינדקס לכרכים א' עד כ' (כולל חוברת 2) עפ"י כותרים.....34

חדשות מקבוצת העניין

עפר דרורי



כח' כסלו' תשע"ג
1 בדצמבר, 2013

"מידע וטקסט"

עלון קבוצת עניין אחזור מידע וטקסט - SIGTRTS
חוברת 2 כרך כ' - דצמבר 2013

חדשות מקבוצת העניין

שלום לכולם!

אנו נפגשים שוב בעלון חדש נוסף. עלון זה סוגר את כרך כ' של הקבוצה ובכך אנו מסיימים את השנה ה- 20 לקיום הקבוצה. אני מאחל לכולנו המשך הנאה, עניין ופעילות ברוכה בתחום אחזור המידע.

ראוי לציין שרוב הנושאים 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.sigtrs.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 לוויני (GPS) כונו אותנו לרחוב פועלי צדק 4, ירושלים. נסיעה טובה!

מאז הוצאת הגיליון האחרון (כרך כ' חוברת מס. 1) ביוני 2013 נפגשה הקבוצה פעמיים.

המפגש הראשון התקיים ביולי 2013, במפגש נשמעו ההרצאות:

א. "מבוא לזיהוי דיבור" ע"י עידן בן שלוש
תחום זיהוי דיבור התחיל כסיוע למוגבלויות שונות של אנשים והתפתח מאוד
נושאים קשורים: Voice Recognition, Speech to text, הודגם שימוש מתסכל של מערכת זו.
הוצגו האפשרויות לשימוש בטכנולוגיה וכן הדגמות שונות לזיהוי ביומטרי קולי
ועוד.
הוצגו כלים וטכנולוגיות המאפשרות לארגונים לנהל את הגישה אליהם
באמצעות טכנולוגיה זו כולל חסימה של מתחזים ועוד.

ב. "טכנולוגיה סמנטית להבנת שפה ע"י שאול שקד"
ההרצאה התמקדה בהבנת הטקסט לא בהיבט של קול (Voice) אלא בהבנה של
השפה. הרעיון הוא הפיכת מידע לא מובנה למידע מובנה תוך חילוף ישויות
וקשרים.
המוצר מתבסס על מוצר איטלקי (Expert System) תוך שילוב מנוע אחזור,
ניתוח פונטי ועוד.
נמסרו דוגמאות לרשת סמנטית בעברית וליישומים נוספים גם בשפות אחרות.
המנוע של המוצר יודע לטפל ברשתות סמנטיות של שפות שונות במקביל. הוצגו
יכולות גרפיות להצגת המידע והקשרים שבו.

המפגש השני התקיים באוקטובר 2013, במפגש נשמעו ההרצאות הבאות:

א. גרפים קונספטואליים (CGS), דרך מעניינת להבנה סמנטית ע"י גיל פוקוס.
לעבודה היו מספר מטרות: סיוע בקשור פרטי מידע ב"ווב" 3 בצורה אוטומטית.
יכולת להבנה חלקית של טקסט ועוד. האלגברה הגרפית יכולה לסייע להבנה
סמנטית של טקסטים שונים תוך יכולת לחבר ביניהם לקבלת תשובה. רכיבים
מהמחקר יושמו ביאחו לעידון פרסום הצמוד לדוא"לים.

ב. פרמטרים המסייעים להצלחת דיבור מוחשב ע"י דורון מודן.
בהרצאה הוצגה הטכנולוגיה והשימושים הנפוצים כיום בדיבור ממוחשב.
הודגמו קטעי "דיבור" שנוצרו מחיבור סינטטי של הקלטות אולפן של עיצורים
ומילים וכו'. מסתבר שכאשר האינטונציה לא טובה היא מאוד מפריעה לאוזן
האנושית. הוצגו המאפיינים המאפשרים לדיבור ממוחשב להישמע אנושי.

מאמרים ועבודות

3.

בחוברת זו לא נמסרו מאמרים.

אני מנצל סעיף זה בבקשה לקבלת עבודות או דוחות שונים שבוצעו במסגרות שונות
(עבודה, אוניברסיטה וכו') העוסקות בנושאי הענין של הקבוצה לצורך פרסומם וכמובן
אני מעודד את כל אחד ואחת מכם לשלוח מאמר מפרי עטו בנושאי הקבוצה.

קריאה מהנה.

חסות .4

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

כללי .5

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

פקס: 02-5688714

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>

מבוא לזיהוי דיבור

עידן בן שלוש

זיהוי דיבור

Automatic Speech Recognition (ASR)

עידן שלוש

הגדרה למערכת לזיהוי דיבור

מערכת זיהוי דיבור או מערכת זיהוי קול היא מערכת ממוחשבת (תוכנה וחומרה או תוכנה בלבד) המסוגלת לפענח דיבור אנושי. הקלט למערכת מגיע ממיקרופון שאליו מדבר המשתמש, משיחת טלפון הנקלטת במערכת וכיוצא בזה והפלט הינו המידע הרלוונטי (המילולי) שזוהה באות הדיבור.

רקע כללי על זיהוי דיבור

תחום זיהוי דיבור על ידי מחשב זוכה לאחרונה להתעניינות רבה, עם השקת שירותים מבוססי ממשק קולי על ידי ענקיות הטכנולוגיה. היעד העיקרי של הטכנולוגיה הוא צמצום הפער בין זיהוי דיבור על ידי אדם לבין זיהוי דיבור על ידי מכונה ולכן מדברים היום בעצם על טכנולוגיה של הבנת דיבור (speech understanding) ולא רק על זיהוי (speech recognition).

חברות הענק בתחום עיבוד שפה טבעית עושות כיום שימוש בטכנולוגיות של עיבוד דיבור כחלק מממשק אדם-מכונה טבעי וחכם. סביר להניח כי התפתחות מגמה זו בעתיד הקרוב תגרום לחדירה משמעותית יותר של טכנולוגיית זיהוי קול בכלל וזיהוי דיבור בפרט לחיי היומיום – בעת הפעלת מכשיר סלולארי, מחשב, קונסולות משחקים ובעת נהיגה. הממשקים בין האדם למכונה נעשים מתוחכמים יותר, וכוללים זיהוי דיבור כמו גם זיהוי תנועות גוף, זיהוי רגשות ומרכיבים נוספים, בהם נעשה שימוש בתקשורת בין אדם לאדם.

התפתחויות בתחום

הסתכלות לאחור על התפתחות טכנולוגיית זיהוי דיבור משקפת תהליך הבשלה איטי. כבר בשנות העשרים של המאה ה-20 יצאו לשוק צעצועים שכללו זיהוי דיבור של מילה בודדת. אולם פריצת הטכנולוגיה חלה כשלושים שנה מאוחר יותר, בשנת 1952. בעשר השנים הבאות כללה הטכנולוגיה זיהוי של ספרות ופונמות. בשנות השבעים של המאה ה-20 התפתחה הטכנולוגיה והגיעה לזיהוי של מילים בודדות ואפילו צירופים קצרים במילון קטן, אך לקראת סוף העשור התבצע מעבר לאופן פעולה בלתי תלוי בדובר, והגדלה של מילוני הזיהוי לאלפי מילים. את הדחיפה המשמעותית אז קיבל התחום בזכות תמיכה של משרד ההגנה האמריקאי, במסגרת תכנית DARPA לעידוד מחקר מתקדם.

השלב הבא בטכנולוגיה חל בשנות השמונים והוא מתאפיין בשימוש במודלים סטטיסטיים בתחום זיהוי דיבור, שאפשר הגדלה משמעותית וגמישות של מילון הזיהוי. בשנות התשעים הושקו מוצרים ושירותים מבוססי זיהוי דיבור בשווקים שונים, בעקבות השימוש במחשוב מתקדם, זמין וזול יותר. בשלב זה הטכנולוגיה הגיעה בשפות מסוימות עד כדי זיהוי דיבור ספונטני במילונים בני עשרות אלפי מילים.

טכנולוגיית זיהוי דיבור

טכנולוגיית זיהוי דיבור היא רב-תחומית, ועושה שימוש בגישות ובתחומים רבים כמו עיבוד אותות דיבור לצורך מיצוי מאפיינים מתוך אות הדיבור, מידול סטטיסטי (לרוב על ידי מודל מרקוב חבוי – HMM) של פונמות, שהן יחידות בסיס לשוניות מהן ניתן להרכיב את כל מילות השפה, תהליכי אימון אוטומטיים (מונחים ולא מונחים) של המודלים הסטטיסטיים מתוך בסיס נתונים בדיבור בגישה מונחית מידע (data-driven), מידול לשוני המתאר את ההקשרים הסמוכים והרחוקים יותר בין מילות השפה (ההסתברות שמילים יופיעו בסמיכות למילים אחרות) ואלגוריתמים של חיפוש לצורך ביצוע חיפוש יעיל ואופטימלי של סדרת המילים בעלת ההסתברות הגבוהה ביותר, בהינתן אות דיבור והמידול האקוסטי והלשוני.

אתגרים

בעוד שניתן לראות כי הטכנולוגיה הגיעה לבשלות המאפשרת שימוש ביישומים מסחריים שונים, הרי שישנם עדיין מספר נושאים מאתגרים, הקשורים בעיקר לחסינות הטכנולוגיה והפיכתה למוצר שימושי באפליקציות שונות, כגון איתור מילות מפתח (keyword spotting).

אחד האתגרים עימם מתמודדת הטכנולוגיה קשור לירידת הביצועים בתנאי רעש: רמת זיהוי הדיבור בסביבות רועשות, כמו רכב נוסע או מקום ציבורי, היא נמוכה יותר, ולעתים קרובות אינה מאפשרת הפעלה של שירות מבוסס זיהוי דיבור.

כמו כן, יש לזכור, כי הטכנולוגיה תלויה בשפה, ובעת המעבר לשפה חדשה נדרשים בסיסי נתונים גדולים בדיבור, לצורך תהליך אימון ייעודי.

יישומים בממשקי אדם-מכונה ובעיבוד דיבור

במעבר מהדיון בטכנולוגיה עצמה לדיון ביישומים האפשריים שלה, ניתן לאפיין שני סוגי יישומים עיקריים לזיהוי דיבור: האחד בא לידי ביטוי בתקשורת אדם-מכונה באמצעות דיבור, והשני בהמרה של תוכן דיבור לטקסט, לצורך עיבוד טקסטואלי.

עם הסוג הראשון נמנית מערכת דיאלוג קולי (spoken dialog system) שבה ניתן לעשות שימוש בזיהוי דיבור בכל מקום בו יש תקשורת אדם-מכונה המתקיימת כיום על ידי הקשה על מקשים. בין השימושים האפשריים לדיאלוג כזה ניתן למנות יישומים כמו מרכזי שירות לקוחות שבהם לפחות חלק מהשיחה תתבצע מול מערכת ממוחשבת. ממשק דיבור אפשרי גם באמצעות המכשיר הסלולארי, כגון הכתבת הודעות טקסט קצרות (SMS) או דואר אלקטרוני, קביעת פגישות, קבלת מידע וחיפוש קולי. בתחום הרפואי יתאפשרו הכתבות של דיאגנוזות רפואיות לתיק הרפואי של החולה, תוך מתן אפשרות לרופא לבדוק את החולה במקביל להכתבת הדיאגנוזה. אפליקציית "הבית החכם" מאפשרת הפעלה קולית של המוצרים בבית: כבר בתקופה הקרובה ניתן יהיה לראות טלוויזיות המופעלות בדיבור, ומשם ניתן יהיה להרחיב את האפליקציה ליתר המכשירים בבית. ומה ברכב? בעתיד הקרוב תתאפשר הפעלה של מערכות לא קריטיות באמצעות דיבור.

הסוג השני מתאפיין בשימוש בזיהוי דיבור בכל מקום שבו יש תוכן קולי או תוכן מולטימדיה, המאפשר לתמלל את פס הקול. המטרה היא לעשות שימוש בכלי עיבוד טקסטואליים כמו אינדוקס וחיפוש על התמליל שנוצר על ידי מנוע זיהוי הדיבור, תוך התחשבות בשגיאות הזיהוי של המנוע.

בין השימושים האפשריים של זיהוי דיבור ניתן למנות גם אפליקציות בתחום המודיעין העסקי – הקלטת השיחות במרכזי שירות לקוחות ותמלולן תאפשר ניתוח טקסטואלי של השיחות לצורך קבלת מידע עסקי מתוכן השיחות, כמו גם אימון ומשוב לנציגי שירות הלקוחות.

לסיכום

מומחים טוענים שמוטיב המפתח בדור הבא של טכנולוגיית זיהוי הדיבור הוא "כמות מובילה לאיכות". שלב זה יתאפיין בשימוש במקורות ידע עצומים ומגוונים, שיכלול, מעבר לשימוש בבסיסי נתונים מקיפים בהרבה מאלה שנעשה בהם שימוש עד היום, גם ידע רב וגמיש על זיהוי דובר, זיהוי רגשות, עיבוד סמנטי ועוד.

שאל שקד

טכנולוגיה סמנטית להבנת שפה באמצעות מנוע לגילוי תובנות

NSDE

(Ness Smart Discovery Engine)

שאל שיקד

רקע

עובד הידע (Information Worker) של ימינו הורגל ע"י Google לראות את העולם דרך מסך החיפוש. הוא מקליד אוסף מילים, והמערכת מחזירת לו תשובות רלוונטיות ממרחבי הרשת. היכולות המדהימות של מנועי החיפוש ברשת גורמת למשתמשים ולאנשי IT לחשוב כי אם רק יבחרו את מנוע החיפוש המתאים - מאגרי המידע האגורים בארגון שלהם יידעו להחזיר תשובה לשאלותיהם. כמעט כל לקוח פוטנציאלי שפגשנו שאל אותנו: "מדוע אני יכול לקבל תשובות כמו מ-Google?". הבעיה של ארגונים (Enterprises) שונה מזו ש-Google פותר:

1. שאלות ארגוניות עוסקות בישויות כמו אנשים (כגון לקוחות), ארגונים/חברות, מוצרים וכו'. המידע על ישויות אלה חבוי באוסף של מסמכים ופרטי מידע מובנים ומובנים למחצה ברחבי הארגון. התשובה לשאלה ארגונית/עסקית נדרשת להצליב בין כלל המידע על כל סוגיו והפורמטים שלו.
 2. קשר בין ישויות הוא מידע בעל ערך עסקי רב. במקרים רבים שאלת המשתמש תהיה על ישויות הקשורות לישויות אחרות (למשל "אדם בשם X המועסק בחברה, העוסקת ב-Y").
 3. לארגונים יש לעיתים מערכות תפעוליות קיימות, העונות לשאלות ייעודיות (לדוגמא: מערכת התאמת טביעת אצבעות, או התאמת אנשים לפי מאפיינים מסוימים). למעשה הצורך הוא בשילוב של מערכות אלה בשאלתה האירגונית.
- לעיתים ארגונים מאמינים כי מנוע אחזור חדש ישפר משמעותית את יכולתם לענות על שאלות ארגוניות. כמוסבר לעיל דרישות האחזור הארגוני רחבות בהרבה מאחזור מסמכים Google style, - ומנוע אחזור הוא רק חלק אחד (אם כי חשוב) בפתרון.
- פתרון לאחזור ארגוני מחייב שילוב של כלים שונים, וכן שילוב של תשתית חזקה אשר תשלב אותם בצורה אינטגרטיבית.
- הניסיון בשילוב של מנועי אחזור במגוון ארגונים מעיד על המורכבות הרבה של עולם האחזור, - נדרשת הבנה עמוקה, וניסיון רלבנטי.
- לצורך זה בנינו את NSDE כתשתית המאפשרת לגייס את כל המידע של הארגון למתן תשובה עסקית/ארגונית.

מהו NSDE?

תחום המודיעין בנס TSG מימש מגוון של מערכות מידע עבור ארגוני מודיעין גדולים במהלך העשור האחרון. אלה מערכות אשר נועדו לממש חזון של מאגר מידע אחיד לכל הארגון. מערכות אלה טיפלו בכל מחזור החיים של מידע מודיעיני, מקליטתו במערכת, הערכתו, ביצוע של קטלוג ומיצוי ידע, ושירותי אחזור מתוככמים.

על בסיס הניסיון שהצטבר, וכן דרישות לקוח חדשות, החלטנו לממש תשתית אשר תספק לארגון (אזרחי או ממשלתי) שירות אחזור אחד, אשר:

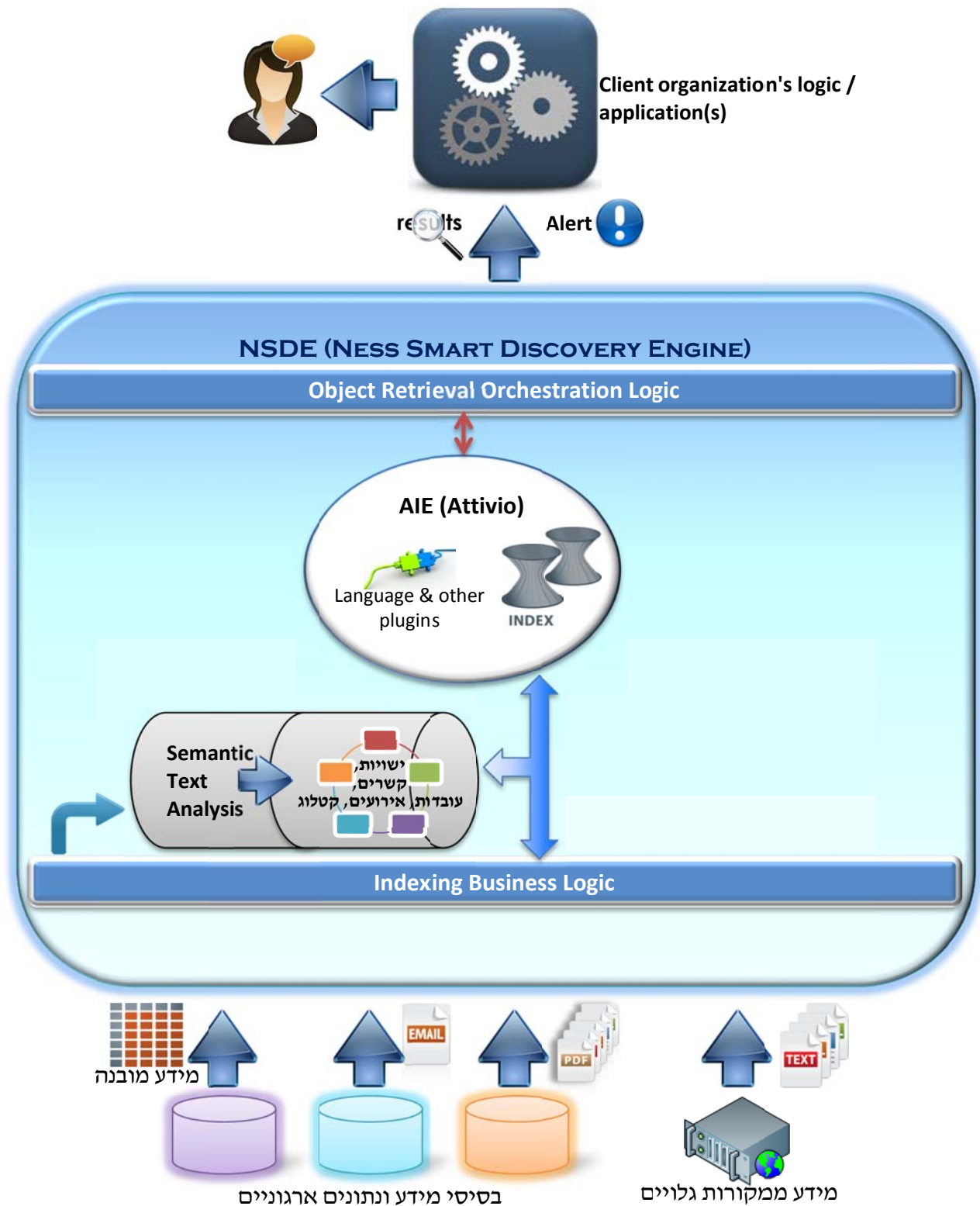
- ✓ אינו כבול למערכת יחידה.
- ✓ יכול לאחד את איי המידע בארגון.
- ✓ מפעיל על כלל המערכות בארגון כלי אחזור ואנליזה שונים.

NSDE הוא התשובה:

- ✓ ניתן להתאמה למערכות שונות, ויכול להתחבר למספר מערכות ומקורות מידע בארגון.
- ✓ אינו חלק ממערכת מסוימת בארגון, אלא מהווה שירות לתשאול והצלבות מידע + מאגר אינדקסים לכלל המידע שבידי הארגון.

תיאור טכני של NSDE

תפיסת הפתרון - ניתוח והצלבה של כלל המידע והידע הארגוני



מודל המידע

כל ארגון מנהל מידע וידע לגבי הישויות בהן הוא עוסק. גם כאשר ארגונים שונים עוסקים לכאורה באותן ישויות מידע - הם מסתכלים עליהן מזוויות שונות, למטרות שונות, ולכן ישויות מאותו סוג נראות אחרת בארגונים שונים. מעבר לכך - ייתכן שבאגפים שונים של אותו ארגון יתייחסו לאותה ישות בצורה שונה, ושהישות תשתנה במשך הזמן.

NSDE מיישם תפיסה גמישה של ישויות מידע. הישות מורכבת מפריטי מידע במבנים שונים - מסמכים, רשומות המגיעות מבסיסי נתונים, ופריטי מידע semi structured, אשר מכילים קטעי טקסט ושדות מובנים. פריט מידע יכול להשתתף במספר ישויות שונות.

דוגמה: פריט מידע, המתאר פגישה עסקית שנערכה בין שלושה משתתפים במשרדי חברה X, מוצא את ביטוי ב:

- ✓ ישויות של שלושת האנשים שנפגשו,
- ✓ בישות של החברה X,
- ✓ ובישות הפגישה עצמה.

NSDE מאפשר שאילתות כלפי מאגר הישויות, המאונדקסות אצלו - שאילתות אשר תצלבנה מידע מכל פריטי המידע העוסקים בישות מסוימת.

הרחבת הדוגמה: נניח שקיים עוד פריט מידע אחר, שבו נאמר כי שני אנשים נוספים נכחו בפגישה. מנגנון האחזור של NSDE משלב את הנתונים מכל פריטי המידע, כך שאם נשאל על הצלבה בין פגישות בין אדם המופיע בפריט הראשון, לאדם המופיע בפריט השני - נקבל את הפגישה האמורה, למרות שאין דיווח אחד אשר אומר כי שני האנשים נכחו באותה פגישה.

מוצרים מסחריים המשולבים ב NSDE

מנוע אחזור Active Intelligence Engine (AIE) של חברת אטיביו (Attivio)

AIE הוא מנוע אחזור מסחרי מוביל, המבוסס על Lucene. ל-AIE מספר תכונות ייחודיות, כמו הגדרת תהליכים גמישה לתהליכים ייעודיים, שהלקוח מבקש לשלב בשלבי האחזור והאינדוקס, יכולת דינאמית לחישוב facets, חילוץ ישויות ועוד. ל-AIE יכולת גידול בלתי מוגבלת, וניתן לבנות מאגר לאחזור של מיליארדי מסמכים, - מאגר שישען על שרתים סטנדרטיים ולא יקרים. AIE מאפשר אחזור על סמך הצלבת תכונות של פריטי מידע - הצלבה זו מבוצעת במהירות גבוהה בסדרי גודל ממהירות ההצלבה המקבילה שבבסיסי נתונים יחסיים מובילים. בנוסף, AIE מאפשר גישת SQL סטנדרטית לבסיס נתונים יחסי. הדבר מאפשר גם לתוכנת בסיס נתונים רגילה לקרוא מתוך מאגר אינדקסים של AIE. דוגמה לשימוש ביכולת זו היא הפקת דוחות ממאגר אינדקסים של AIE באמצעות מחוללי דוחות מסחריים סטנדרטיים.

ניתוח סמנטי של טקסט

נס TSG פתחה גרסה עברית של מנוע ניתוח סמנטי אמתי. הפיתוח מבוסס על פלטפורמה קיימת, שנס התמחתה בה, בה קיים כבר מימוש מסחרי עבור מספר שפות אחרות - כמו אנגלית וערבית.

הניתוח הסמנטי מבצע ניתוחי מתקדם (ניתוח מורפולוגי, תחבירי וכן ניתוח המשמעות), והוא מספק (על סמך אונטולוגיה שהלקוח יכול להתאים לצרכיו) :

- ✓ חילוץ ישויות.
- ✓ חילוץ קשרים בין הישויות.
- ✓ חילוץ עובדות ואירועים.
- ✓ חילוץ של מאפייני זמן ומקום.
- ✓ קטלוג התוכן לעולם הקטגוריות של הלקוח (או לעולם קטגוריות כללי).
- ✓ Sentiment Analysis.

גרפים קונספטואליים (CGS) דרך מעניינת להבנה סמנטית

ד"ר גיל פוקס

גרפים קונספטואליים (CGS), דרך מעניינת להבנה סמנטית

גיל פוקס

במאמר זה מוצגת מערכת דדוקטיבית עם ייצוג מובנה אשר בנויה מגרפים רעיוניים משופרת בהיררכיית סוגים שמבצעת התאמה ואחזור באופן ליניארי ואסוציאטיבי המשתמש ברשומות באורך משתנה.

הטכניקה של המערכת הזו כרוכה בתרגום של קטעי טקסט לתוך פורמט מסודר אינפורמטיבית (בייצוג גרף סמנטי של ידע) ושליפה מתוך הסדר ההיררכי המושרה. יישום זה הוא מערכת של גרפים רעיוניים ליניאריים, עיבוד שפה טבעית, התאמת ואחזור שאילתות. שיח אנגלי מומר לגרפים רעיוניים המאוחסנים במסד נתונים אסוציאטיבי. מסד הנתונים ההיררכי מונחה על ידי מורפולוגיות התוכן, היררכיית סוגים אונטולוגית אסוציאטיבית, כדי לייצר סדר חלקי המבוסס על היררכיית "יותר כללי מאשר" המבוססת על שיוך של תת-גרפים.

בשל האופי של מסד נתונים האסוציאטיביים השאילתה נענית על ידי כך שהי מוכנסת לתוך מסד הנתונים. יורשיה של השאילתה במסד הנתונים הם התשובות לשאילתה כמו גם אינפורמציה נשלפת רלבנטית נוספת (וספציפית יותר).

THE CG MARS LANDER

Gil Fuchs¹ and Robert Levinson¹

¹ Department of Computer Science, University of California, Santa Cruz, CA 95060

² in collaboration with TextWise, Syracuse, New York

E-mail: {gil,levinson}@cse.ucsc.edu

Phone: 408-459-2087

Abstract. The name of the system is “THE CG MARS LANDER”, which stands for **T**ype **H**ierarchy **E**nhanced **C**G **M**atching **A**nd **R**etrieval **S**ystem **L**inear **A**ssociative **N**-tuple **D**eductive **E**mbodied **R**epresentation. The application is text retrieval. The technique involves translation of text fragments into an informationally ordered form (CGs) and retrieval on the induced subsumption ordering. This application is a linearized CG, NLP, query, matching and retrieving system. English discourse is converted into CGs which are stored in an associative database. The hierarchical database is guided by content morphology and associated ontological type hierarchy, to produce a partially ordered “more general than” hierarchy based on subgraph containment. Due to the nature of the ADB a query is answered by being inserted into the DB. The query’s successors in the DB are the answers to the query, as well as the (more specific) retrieved relevancies.

1 Introduction

1.1 Background and motivation

CGs have a great promise and many applications, natural language processing and understanding, knowledge aquisition, modeling, data mining to name a few. Development and research in this domain, however, has suffered from the belief that this promising technology will not scale, even on multi processor computers. The initial timings of this project gave comfort to the sceptics. The first attempt (early 1996) at converting 100,000 English sentences to CGs and matching the CG database against a small set of queries took about 6.5 hours on a Sun Sparc Ultra enterprise 4000. Using a variety of techniques, including the progressive application of the Levinson associative database methods III through VI [8], we managed to reduce this to 19 seconds (of which 16 seconds is the overhead ontology loading and 3 seconds is the actual processing time) on the same Sparc station (early 1997). Our goal is to further speed the process up so that this technology can actually be useful on real world size problems.

2 Architecture of the System

2.1 High level architecture

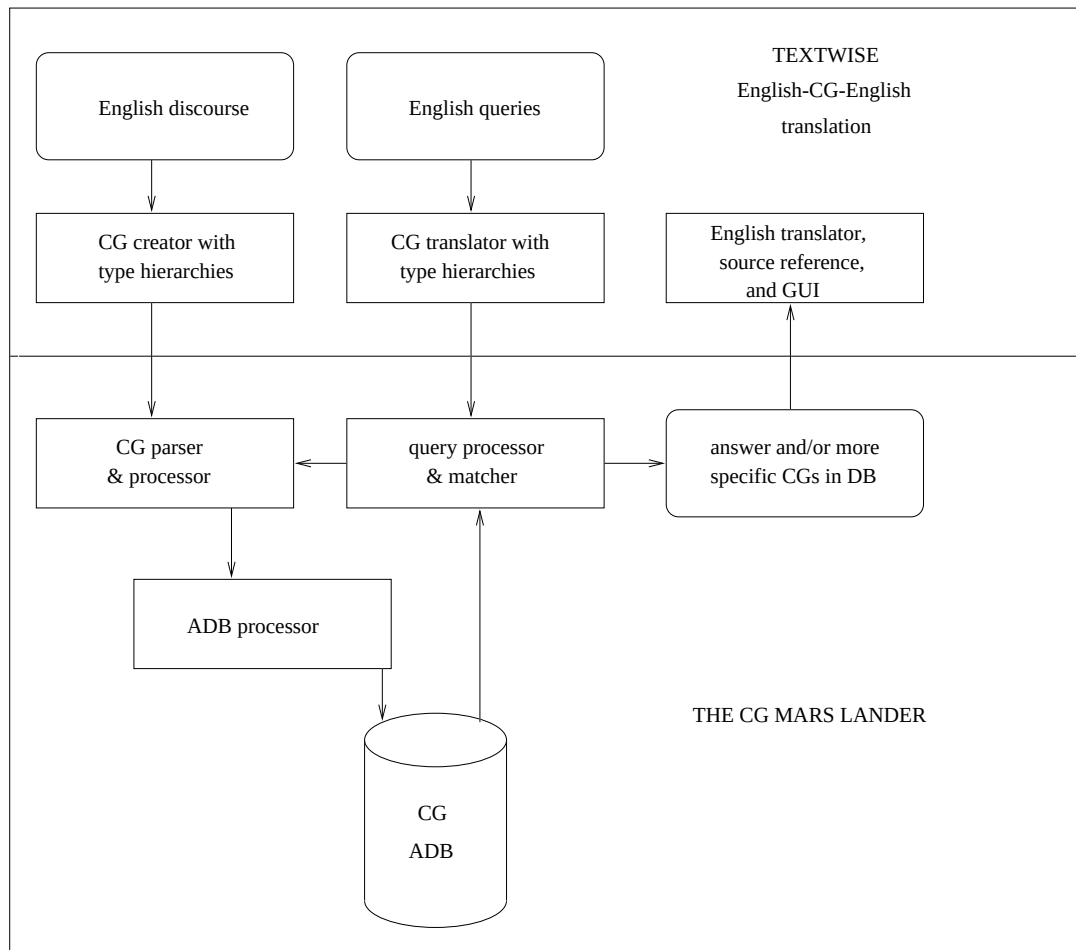


Fig. 1. The CG Mars Lander: Schematic View of Module Architecture

Figure 1 shows the overview of the CG tool. English discourse, coupled with ontology and queries are inputs to the system, and the result is an English answer with appropriate references to where in the English discourse the query gets answered.

2.2 Key features

Simple representation of CGs: By employing a linear tuple notation we provide for a simple and efficient implementation of CGs. A CG consists of n tuples, each of which, has any number of arguments, which, in turn, can either be a linguistic construct or a CG.

Expressiveness of CGs:

- **item negation:** Any CG, relation, or concept node arguments may be negated.
- **argument values:** Any concept node argument may have a value.
- **ranges:** Any place a single value is allowed, a range is also permissible.
- **dates:** These ranges, may be of **date** or **time** type.

Arbitrary nesting capability of CGs: An argument can, in turn, be a previously introduced CG tag. Hence, allowing nested contexts to any depth.

Methods III through VI of associative DB matching: Exploiting Levinson's algorithms for optimized graph retrieval.

Automatic maintenance of the type hierarchy: The type hierarchy is an **isa** partial order over some 100,000 English concepts.

Automatic cycle recognition for the type hierarchy: Should the above hierarchy contain cycles, a warning is issued.

DSPS codes: Each CG may contain an association to the document, section, paragraph and sentence from which it was derived, for bibliographic retrieval.

2.3 Levinson's UDS [8] design foundations slogans

- Every primitive data object, label or symbol should be stored only once with pointers used to denote the actual uses of the object.
- Every compound object should be stored with the minimum information required to represent the combination of its parts.
- Given no loss of accuracy, objects should be processed at the highest level of abstraction possible.
- If one were to implement a conceptual graph based on the diagrammatic representation, the costs associated with storage and matching would be much higher than they need to be.
- The same abstraction mechanism that goes from labels to graphs can be taken one step further to facilitate the storage and retrieval of nested context graphs.
- A graph is itself the best descriptor of its nodes.

3 Application of the System

3.1 Summary of results and timings

Our current program can accept a large set of CGs (tens of thousands), an ontology of some few hundred thousands words, and a set of queries. The program stores the CGs in a database which can be saved and restored, and answer queries, by returning relevant CGs from the previously saved DB. The timing statistics on a Sun Ultra Enterprise 4000 (with 4 UltraSPARC 167Mhz and 512KB External Cache CPU and 256BM of main memory) are as follows. Read, process, and store an 18,000 CG input file in 1 hour and 46 minutes. Reloading of above DB takes on the order of seconds. A 150,000 word ontology is processed in 16 seconds. Each query is handled in 5.5 seconds. For smaller databases (hundreds of CGs only), the time to handle a single query can be as low as 0.2

seconds. A typical CG consists of some ten tuples, each of which has two to five arguments. Some large CGs can, however, reach up to 30 tuples (with no effective limit in the program). CG processing includes the treatment of entity negation, item values and ranges (including dates), as well as nesting.

3.2 Sample session

Samples of elements in the package:

- An English sentence:

```
The US government's privately managed Hungarian American
Enterprise Fund has invested 1 million Dollars in
Hungary's first business to establish automated teller
machines nationwide.
```

- Tokenization and tagging using TextWise's parser and conversion to CGs. One sentence may give rise to several CGs, with embedding also possible.
- Standard CG notation:

```
[[government]-->(AGNT) --> [be]] --> (AGNT) --> manage
--> (OBJ)
--->
[[Hungarian_American_Enterprise_Fund] --> (AGNT)
--> [invest] --> (OBJ) --> [dollars: 1000000]
--> (IN) --> [First-Business]].
```

Note that the sentence parsed above is not related to the question and answer which follow.

- Sample queries:

```
/* Q1: When did Rupert Murdoch own the New York Post? */
[[Rupert Murdoch] -->(AGNT) --> [own] --> (OBJ)
---> [New_York_Post]] --> (PTIM) --> [???].
```

- Answer:

```
/* A1: Rupert Murdoch owned the New York Post from
1976 to 1988.
*/
[[Rupert Murdoch] -->(AGNT) --> [own]
--> (OBJ) --> [New_York_Post]]
--> (PTIM) --> [year: (1976,1988)].
```

3.3 Cost benefit analysis

Suppose one only plans to do a small number of queries Q , over a database of N CGs. The question arises whether it is worth the overhead of creating a

Levinson Method III database, in which retrievals and insertions take approximately $\log_{10}^2(N)$ comparisons for a database of size N . The alternative is simply to compare each query to every CG.

Here we do the necessary math to aid in the decision: No precompilation of database: cost is $N * Q$ graph comparisons. (call this Method I)

Creating method III database: average cost per insertion is approximately $\log_{10}^2(\frac{N}{2})$, giving a cost to create the database of $N \log_{10}^2(\frac{N}{2})$ and a cost to answer the queries of $Q \log_{10}^2(N)$.

So the question is when NQ is $< (N \log_{10}^2(\frac{N}{2}) + Q \log_{10}^2(N))$.

Here is a table of results:

N	Q	Method I Cost	Method III Cost
10	1	10	5.00
10	10	100	14.90
10	100	1,000	104.80
100	1	100	296.60
100	10	1,000	328.64
100	100	10,000	688.60
1,000	1	1,000	7,293.43
1,000	10	10,000	7,374.44
1,000	100	100,000	8,184.40
1,000	1,000	1,000,000	16,284.44
10,000	1,000	10,000,000	152,823.78
10,000	10,000	100,000,000	296,823.78

Table 1. Comparison table between costs of respective methods

From this it can clearly be seen, that Method III is most cost effective except for very small ratios of queries to database size, and for very high ratios the savings grow exponentially.

Levinson's Methods IV-VI provide further order of magnitude improvements over these numbers.

4 Limitations of the System

4.1 Open questions and concerns

The translation from an English sentence to a CG is not, by all means, unique. One illustrative example might be the following sentence.

Rupert Murdoch owns Fox.

Which can be rendered either as a CG showing a characteristic (possession) or as a standard agent-verb-object relation.

[Rupert Murdoch] --> (POSS) --> [Fox].

[Rupert Murdoch] --> (AGNT) --> [own] --> (OBJ) --> [Fox].

Apart from the need for consistency, it isn't clear which of the above alternatives is better and why. Do we need to add graph transformation rules to exploit these equivalencies (in addition to the type hierarchy)? Should we create a mechanism for inferencing? At this time, these are still open questions.

5 Acknowledgements

The work on this project was done in collaboration with the research staff of TextWise, Syracuse, New York. TextWise and Manning and Napier Information Systems have provided the entire financial support for the project. Some of the work is part of a larger SBIR commercialization project at TextWise administered by Rome Labs. All commercial rights are the property of TextWise and Manning and Napier Information Systems.

Relevant papers on which this work is based [1, 2, 3, 4, 5, 6, 7, 8] are included in the list of references that follows:

References

1. C. Colin and R. Levinson, "Partial order maintenance," *Special Interest Group on Information Retrieval Forum*, vol. 23, no. 3,4, pp. 34–59, 1988.
2. G. Ellis, R. A. Levinson, and P. Robinson, "Managing complex objects in PEIRCE," *Special Issue on Object-Oriented Approaches in Artificial Intelligence and Human-Computer Interaction (IJMMS)*, vol. 41, pp. 109–148, 1994.
3. R. Hughey, R. Levinson, and J. D. Roberts, eds., *Issues in Parallel Hardware for Graph Retrieval*, 1993.
4. R. Levinson, "A self-organizing retrieval system for graphs," in *AAAI-84*, pp. 203–206, Morgan Kaufman, 1984.
5. R. Levinson, "Pattern associativity and the retrieval of semantic networks," *Computers and Mathematics with Applications*, vol. 23, no. 6-9, pp. 573–600, 1992. Part 2 of Special Issue on Semantic Networks in Artificial Intelligence, Fritz Lehmann, editor. Also reprinted on pages 573–600 of the book, *Semantic Networks in Artificial Intelligence*, Fritz Lehmann, editor, Pergammon Press, 1992.
6. R. Levinson and G. Ellis, "Multilevel hierarchical retrieval," *Knowledge-Based Systems*, vol. 5, pp. 233–244, September 1992. Special Issue on Conceptual Graphs.
7. R. Levinson and G. Fuchs, "A pattern-weight formulation of search knowledge," Tech. Rep. UCSC-CRL-91-15, University of California Santa Cruz, 1994a. Revision to appear in *Computational Intelligence*.
8. R. A. Levinson, "UDS: A universal data structure," in *Proc. 2nd International Conference on Conceptual Structures*, (College Park, Maryland USA), pp. 230–250, 1991.

This article was processed using the $\text{\LaTeX}$ macro package with LLNCS style

פרמטרים המסייעים להצלחת דיבור מוחשב

דורון מודן

פרמטרים המסייעים להצלחת דיבור ממוחשב

דורון מודן

את הדיבור הממוחשב, ה-TTS (Text to Speech) אין כמעט צורך להציג כיום. הוא משמש בחיי היום-יום, למשל במקרי לקות ראייה, לקות קריאה, מוגבלות בדיבור, משחרר את הידיים והעיניים ('קריאה' בזמן נהיגה), שילוב עם מערכת זיהוי דיבור (בעיקר בטלפוניה) ושימור הקול לקראת ניתוח להסרת שפתות הקול.

בהצגת דיבור מבחינה אקוסטית, ניתן להסתפק, בדומה למוזיקה, בארבעה פרמטרים בלבד: גוון (בדומה לגוון של כלי נגינה, מדובר בתגובתו של קול ראשוני למסלול תהודה, מלבד זה שבמקום צלילים, אנו מפרשים את התוצאה כהגאים), וכן גובה טון, משך ועוצמה.

האתגר העיקרי העומד בפני מערכת לדיבור ממוחשב הוא כפול:

1. ניתוח נכון של המבנים ומציאת תבניות של ארבעת הפרמטרים הנ"ל.
2. בחירת יחידות לשרשור, כך שיתאימו לתבניות שנקבעו, מבלי שההתאמה הזו תישמע מאולצת.

מאז שהשיטה השלטת בדיבור ממוחשב היא של שרשור מתוחכם של מקטעי הקלטות מתוך קורפוס מוקלט מראש, נעשה קל מאוד להפוך כמויות גדולות של טקסט לכדי דיבור. זאת, כמובן, לאחר שנעשה ניתוח אוטומטי של הטקסט (למשל, TTP, דהיינו Text to Phoneme, המרת טקסט רגיל להגאים), ניתוח תחבירי על גזרותיו הפרוזודיות, נורמליזציה של הטקסט (למשל התמודדות עם סימנים, ספרות וכו' והפיכתם להגאים) ועוד.

יחד עם זאת, ברור שכל שרשור של הקלטות הוא בהגדרתו איננו דיבור טבעי, ומטרתה של מערכת טובה היא 'לנהל את הסיכונים' ולמצוא את השרשור שהוא הרע במיעוטו, בשאיפה שאי-הטבעיות הקיימת תהיה כה זניחה, עד שאוזן אנושית לא תבחין בה.

כמו בביצוע מוזיקה, כך גם בדיבור, קיימות תבניות בלתי מודעות של גוון, גובה טון משך ועוצמה. אם ננהג בהתאם לתבניות טבעיות אלה, הדיבור יישמע טבעי. אחרת, כתוצאה משרשור של מקטעים ש'לא מתחברים יפה' נשמע תופעות של רעידה, "glitch", 'חורים' ושלל מוזרויות, המאפיינות לעיתים את הדיבור הממוחשב. ניתוח טוב, יחד עם שרשור זהיר, יכול לגרום לכך שדיבור ממוחשב לא יישמע שונה בהרבה מדיבור טבעי, ונראה, כי הטכנולוגיה צועדת לכיוון זה.

לדוגמה, פרמטר המשך בדיבור מושפע מגורמים רבים, במספר רמות הלשון. למשל: לכל ההגה יש משך "פנימי", ארוך יותר או קצר יותר בהשוואה לאחרים, הנקבע בדרך כלל מסיבה פיזיולוגית. משכו של ההגה נקבע גם ברמת ההברה, מזהות ההגה השכן. ידוע גם שהברה מוטעמת משכה גדול יותר. כמו כן, מילת הגרעין במשפט נוטה להיות ארוכה הרבה יותר מחברותיה. יתר על כן, ניתוח תחבירי (parsing) אוטומטי יגלה לנו שיש יחידות במשפט,

המקושרות חזק יותר זו לזו, בקבוצה שונה מיחידות אחרות במשפט. מתגלה, כי בדיבור טבעי יש נטייה להאט לפני גבול קבוצה תחבירית, ונטייה להאיץ מיד ביציאה מגבול זה. חיקוי כל הגורמים האלה, ואחרים, על ידי המערכת האוטומטית הוא הכרחי ליצירת דיבור מובן ואפקט טבעי.

גם גובה הטון נקבע בידי תבניות ידועות מראש, והוא גורם מכריע להצלחה של התאמת מקטע מוקלט לשרשור. זאת משום שקל יותר "לרמות את האוזן" ולגרום לה להאמין שמשך או עצמה ששונה באורח מלאכותי הם טבעיים, מאשר להאמין לגובה טון שהשתנה בצורה ניכרת.

לסיכום, את הקורפוס המוקלט מראש לא ניתן לשנות אבל לנתן לעדן את הניתוח ביחס למשך, וכן ביחס לעצמה ולגובה הטון (בדרך כלל יש קורלציה בין שלושת הפרמטרים האלה) והדבר יכול לתרום מאוד לתרום לבחירת היחידות לשרשור, לצורך הפחתת הזרות לאוזן האנושית.

אינדקס לכרכים א' עד כ' (כולל חוברת 2) עפ"י מחברים

אינדקס לפי שם מחבר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
XML והשלכותיו על בסיסי נתונים ואחזור טקסט	אבגי ראובן	ו'	2	6/2000
מנוע האחזור Inter Text - גרסה חדשה	אבגי ראובן	ט'	1	1/2003
תכונות מנוע החיפוש Inter Text	אבגי ראובן	ט'	1	1/2003
החפשו של סנונית	אבולוף אוריאל	ה'	2	5/1999
טיפול בנושא השמות ביד ושם	אברהם אלכס	ט'	2	6/2003
Yad Vashem names and places index	אברהם אלכס	ט'	2	6/2003
חיפוש מידע ברשת	אדלשטיין קובי טל רפפורט	ח'	1	1/2002
הטכנולוגיה הישראלית שמאפשרת להרכיב את הפאזל היהודי	אדרת עופר	יט'	1	6/2012
מנוע חיפוש וניהול ידע בעברית	אופיר עידית זהבי יורם	ח'	2	6/2002
המרת שאילתות בשפה טבעית לשאילתות SQL	אופק נועה עפרי דר	יא'	1	1/2005
Full Tset - מנוע חיפוש בעברית	אורון שחר	ז'	1	1/2001
מנוע חיפוש בטקסטים עבריים	אורנן עוזי	יא'	2	6/2005
תכונות מנוע האחזור - ניאוזאורוס	אורנן עוזי	יב'	1	1/2006
ניאוזאורוס - מנוע חיפוש וניהול מימדים	אורנן עוזי	יב'	1	1/2006
ממנוע חיפוש לשלטי דרכים	אורנן עוזי	יג'	1	1/2007
מבנה המילה והשתקפותו בניקוד ובתעתיק	פרופ' אורנן עוזי	יג'	2	6/2007
GUIDance : כלי ליצירת ממשק גרפי למערכת MF	אורנשטיין דרור	א'	2	10/1995
בחינת מודל ללמידה דרך היפרטקסט	אחיטוב שמיר	ב'	1	5/1996
ארכיון צה"ל ומערכת הבטחון	אלגום אורי	ב'	1	5/1996
שרות החיפוש של גוגל	אליאסי אמיר	ב'	2	6/2009
כלים לביצוע מחקר איטרטיבי	אמיתי יעל	ג'	1	2/1997
מנוע החיפוש RetrievalWare	אנדלמן נחמה צבי קמר	ט'	2	6/2003
הגניזה הקהירית ממשיכה לספק כותרות	דר' אפשטיין יכן	יט'	1	6/2012
ייצור אוטומטי של תיאורי ומילונים דו לשוניים	ארד איריס	י'	2	6/2004
מטריקה, סגמנטציה ומבנה קואורדינטות במערכת אחזור טקסט	ארד איריס	יב'	1	1/2006
שימוש באונטולוגיות לניהול וארגון מידע	ארד רינה	יג'	1	1/2007
פיתוח מערכת טקסטואלית תומכת החלטה בסביבה מרובת פלטפורמות	אריאלי אהוד	ז'	1	1/2001
חיפוש ארגוני, סיכום מפגש שולחן עגול	בודוגין ליוז	יט'	2	12/2012
ניהול ידע בחקירות	בלומקין נמרוד	יח'	2	1/2012
השוואת מערכות חיפוש של המאגר ה-Medline	בנחקן אלינור	ה'	2	5/1999
חבילת מוצרים לניתוח ואחזור שמות	בן אהרון דביר	ט'	2	6/2003
מבוא לזיהוי דיבור	בן שלוש עידן	כ'	2	10/2013
יציבות מידע ברשת	דר' בר-אילן יהודית	ו'	2	6/2000
ארכיון הכתבות הדיגיטליות בידיעות אחרונות	ברדוש יוסי	י'	1	1/2004
אחזור קבצי מוזיקה ואודיו באינטרנט	בר סימן טוב פני	טו'	1	1/2009
ממשק שפה טבעית בעברית למסכי נתונים יחסיים	גור אלי	א'	2	10/1995
היפרטקסט וקבלת החלטות	גוראון רן	ד'	1	2/1998
חיפוש חופשי בטקסטים סרוקים בעברית	גורא שמעוני	יא'	1	1/2005
מורפולוגיה למנועי חיפוש	גורא שמעוני	יד'	1	1/2008
ישום GSA במאגר פסקי דין	גורא שמעוני	יט'	2	12/2012
השוואה של מנועי חיפוש בעברית	גיל תומר	ח'	2	6/2002
פתרונות לאבטחת איכות תוכנה	גילעד זוהר	ד'	2	5/1998
"שימוש בטכנולוגיות מבוססות קוד-פתוח ביישומי אחזור-טקסט"	גליבוב ליאוניד וויסיפון אורן	יז'	1	1/2011
טכנולוגיית ה-Push	גליקשטיין טליה	ג'	2	7/1997
מיומנו של מחפש עצמאי - איך מחפשים ומוצאים	גרינברג זאב	טו'	1	1/2009
RexyGo מנוע חיפוש חדש	דבש יוסי	טז'	1	1/2010

המשך אינדקס לפי שם מחבר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
הבנת שפה ביישומי טקסטים	ד"ר דגן עידו	יג'	2	6/2007
טכנולוגיית מיקרוסופט לאחזור מידע (ביג דטה)	דוד מאור	כ'	1	6/2013
הוצאה לאור אלקטרונית	דנברג אמיר	ג'	1	2/1997
הסבת ארכיון צה"ל ממחשב WANG למחשב HP תחת UNIX	דנברג בני	ד'	1	2/1998
חדשות מקבוצת העניין	דרורי עפר	א'	1	4/1995
		א'	2	10/1995
		ב'	1	5/1996
		ב'	2	11/1996
		ג'	1	2/1997
		ג'	2	7/1997
		ד'	1	3/1998
		ד'	2	5/1998
		ה'	1	1/1999
		ה'	2	5/1999
		ו'	1	1/2000
		ו'	2	6/2000
		ז'	1	1/2001
		ז'	2	6/2001
		ח'	1	1/2002
		ח'	2	6/2002
		ט'	1	1/2003
		ט'	2	6/2003
		י'	1	1/2004
		י'	2	6/2004
		יא'	1	1/2005
		יא'	2	6/2005
		יב'	1	1/2006
		יב'	2	6/2006
		יג'	1	1/2007
		יג'	2	6/2007
		יד'	1	1/2008
		יד'	2	6/2008
		טו'	1	1/2009
		טו'	2	6/2009
		טז'	1	1/2010
		טז'	2	6/2010
		יז'	1	1/2011
		יח'	1	7/2011
		יח'	2	1/2012
		יט'	1	6/2012
		יט'	2	12/2012
		כ'	1	6/2013
		כ'	2	10/2013
איך לצמצם את שטח האחסון בדואר גיימייל	דרורי עפר	יח'	1	7/2011
כיצד לייצר ספר אלקטרוני עם תכונות חיפוש באמצעות Adobe Acrobat	דרורי עפר	יא'	1	1/2005
הצגת תוצאות חיפוש במערכות אחזור מידע תוך שימוש באלמנטים שונים לייצוג המסמכים - מחקר שימוש	דרורי עפר	יא'	1	1/2005
בניית מאגרי-מידע גדולים	דרורי עפר	ה'	2	5/1999
ישום מערכת איחזור מידע טקסטואלי בשע"ם	דרורי עפר	א'	1	4/1995
מנועי חיפוש באינטרנט	דרורי עפר	ג'	1	2/1997
ניהול וארגון קבוצת עניין	דרורי עפר	ב'	1	5/1996

המשך אינדקס לפי שם מחבר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
עיצוב ממשק משתמש במערכות מידע	דרורי עפר	ה'	1	1/1999
שילוב מערכות אחזור טקסט ומערכות מידע קובנציונליות	דרורי עפר	ג'	2	7/1997
הצגת תוצאת חיפוש בממשק משתמש במערכות אחזור טקסט - סקירת ספרות	דרורי עפר	ו'	1	1/2000
הוספת מנוע אחזור לאתר אינטרנט	דרורי עפר	ן'	2	6/2000
קריטריונים להשוואה בין מנועי חיפוש	דרורי עפר	ז'	1	1/2001
שילוב בסיסי נתונים ומאגרי-מידע באתר ה- WEB בספריה ובמרכזי מידע	דרורי עפר	ז'	2	6/2001
שילוב בסיסי נתונים ומאגרי-מידע באתר ה- WEB בספריה ובמרכזי מידע	דרורי עפר	ח'	1	1/2002
מנועי חיפוש בעברית - רשימת ספקים	דרורי עפר	ח'	1	6/2002
רשימת ספקים של מנועי אחזור בעברית (11.2002)	דרורי עפר	ט'	1	1/2003
שימוש במילים נפוצות במסמך לאיתור נושא המסמך	דרורי עפר	ט'	1	1/2003
ניהול תוכן - תכונות נדרשות	דרורי עפר	ט'	1	1/2003
קריטריונים לבחירת מנוע אחזור טקסט - גירסה 2	דרורי עפר	ט'	1	1/2003
מנוע חיפוש לשפה העברית	דרורי עפר	ט'	2	6/2003
רשימת ספקים למנועי אחזור בעברית (גירסה 3.2003)	דרורי עפר	ט	2	6/2003
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 3.2004)	דרורי עפר	י'	2	6/2004
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 4.2005)	דרורי עפר	יא'	2	6/2005
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 06.2006)	דרורי עפר	יג'	1	1/2007
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 07.2009)	דרורי עפר	טז'	1	1/2010
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 10.2010)	דרורי עפר	יז'	1	1/2011
קריטריונים לבחירת מנוע אחזור טקסט - גירסה 3	דרורי עפר	י'	2	6/2004
איתור נושא מסמך בצורה אוטומטית תוך שימוש במילים נפוצות	דרורי עפר	י'	2	6/2004
שיקולים בבחירת מערכת לניהול תוכן	דרורי עפר	יא'	2	6/2005
קריטריונים לבחירת מנוע אחזור טקסט - גירסה 4	דרורי עפר	יב'	1	1/2006
דירוג רשימת תוצאות חיפוש - ממצאי מחקר	דרורי עפר	יב'	2	6/2006
שרות אחזור טקסט של גוגל	דרורי עפר	טו'	1	1/2009
תכונות מנוע החיפוש ATTIVIO	דרורי עפר	טז'	1	1/2010
תכונות מנוע החיפוש REXYGO	דרורי עפר	טז'	1	1/2010
תכונות מנוע החיפוש FAST	דרורי עפר	טז'	1	1/2010
תכונות מנוע החיפוש XRS	דרורי עפר	טז'	2	6/2010
ניהול ידע מחקר הצל"שים באתר הגבורה	דרורי עפר	כ'	1	6/2013
Inter Text	הוך איציק	ד'	1	2/1998
מערכת נוהלים בטכנולוגיית אינטרנט	הוך איציק	ד'	2	5/1998
סיכום מצב קיים במערכות אחזור טקסט	דר' הנדזל רות	ב'	2	11/1996
חברי הכנסת כצרכני מידע	דר' מרקוס רבקה	יט'	2	12/2012
עיבוד שפה טבעית בערבית	הרשברג נמרוד כפיר בר	יד'	1	1/2008
עיצוב ממשק משתמש ל- WEB	ווינדפלד צביקה	ד'	2	5/1998
כיווני פיתוח באחזור טקסט TQL	וויזל גלעד	א'	1	4/1995
חיפוש מידע מבוסס ניתוח טקסט	וינהבר אורי	יב'	1	1/2006
ניתוח שאליות באתרי מכירות	דר' ולן אינגריד	יג'	2	6/2007
מחשוב המתווה הלקסיקוגרפי של השפה העברית בת זמננו	זיסמן בן-דור שירה	יא'	2	6/2005
ממשקים ויזואלים לתוצאות חיפוש	זמיר אורן עציוני אורן	ו'	1	1/2000

המשך אינדקס לפי שם מחבר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
ויזואליזציה של תוצאות חיפוש במערכות אחזור מידע	זמיר אורן	ז'	2	6/2001
Advisor - כלי ליצירת יישומי שפה טבעית	חברת סלסנס	טו'	2	6/2009
כלי חיפוש סטנדרטיים במערכות ומוצרים קיימים	חגיז מוטי	טו'	2	6/2006
Web 3.0 מעבר לפינה - טכנולוגיות סמנטיות באינטרנט ובארגונים	חזקיה רוני	טז'	2	6/2010
מגמות עתידיות בעולם איחזור המידע expetext	ד"ר חנני אורי	א'	1	4/1995
מנתונים לידע - MindCite	ד"ר חנני אורי	טו'	2	6/2003
זיהוי תמידי של מידע ברשת האינטרנט	חן בועז	ו'	1	1/2000
ויזואליזציה של מידע בהתבסס על קשרים מובנים וניתוחם	חרדק פבל	טז'	2	6/2010
יין ואינטרנט	טיוטו מרק	ו'	2	6/2000
"הרתחת אוקיינוס האינטרנט": הדור הבא של רשתות חברתיות	ד"ר יהודה יאיר	טז'	2	6/2010
"שימוש בטכנולוגיות מבוססות קוד-פתוח ביישומי אחזור-טקסט	יוסיפון אורן	יז'	1	1/2011
השוואת מנועי חיפוש ומנוע ATTIVIO	יעקובוביץ צחי	טו'	2	6/2009
מערכת לאיתור ישויות	יפת אביבה דרורי עפר	י'	1	1/2004
מחקר בתיק דיגיטאלי בבתי המשפט, כיצד?	ירדני ירדן	יח'	1	7/2011
הטכנולוגיה של אוליב ל- OCR ושימוש ב- Fast לאחזור מסמכי בתי משפט	ירדני ירדן	יח'	2	1/2012
תכונות מנוע החיפוש מורפיקס	ירדני לאורה	טו'	1	1/2003
חיפוש טקסט מלא בעברית ובערבית - הבעיה והפתרון	ירדני לאורה	י'	1	1/2004
תכונות מוצר לניהול הידע D2K.NET	כהן אייל	טו'	1	1/2003
חיפוש תלוי הקשר, עקרונות ודוגמאות	כהן חנן	ח'	1	1/2002
זכרון לטווח רחוק	כהן שגיא	יט'	2	12/2012
אחזור מסמכים משובשים	לבנה משה	ח'	1	1/2002
MKknowledge - תוכנה לקבלה, ניתוח והצגה של ידע	לוי ניר	יז'	1	1/2011
מפת הדרכים של החיפוש הארגוני במיקרוסופט- MOSS	לוסטיג רונה	טז'	1	1/2010
הנגשת ארכיונים לציבור הרחב באמצעות דוגמאות מיד ושם	ליבר מיכאל	טו'	2	6/2009
"זה נשמע סינתי! כלים לניתוח טקסט בעברית ובשפות אחרות"	ליטוב דוד אנדרו גולדמן	יז'	1	1/2011
המרת אתרים מעברית וויזואלית ללוגית	ליס אורלי	טו'	2	6/2003
ארגון תוצאות חיפוש ממנועי אחזור באינטרנט	דר' לסט מרק	ח'	2	6/2002
סיווג אוטומטי של מסמכי טקסט בשפות שונות (כולל ערבית)	דר' לסט מרק	יג'	1	1/2007
ממשק חלונאי אחד לטקסטים בסביבות עבודה שונות	מבשב ישראל טל כוכבה	ז'	1	1/2001
מבוא לדיבור ממוחשב, תרגום מכונה וזיהוי דיבור	מודן דורון	יז'	1	1/2011
פרמטרים המסייעים להצלחת דיבור מוחשב	מודן דורון	כ'	2	10/2013
תכונות מנוע החיפוש WizDoc	מידן אברהם	טו'	1	1/2003
WizDoc - מנוע חיפוש לפי משמעויות בעברית ובאנגלית	מידן אברהם	טו'	2	6/2003
חיפוש לפי משמעויות	דר' מידן אברהם	יד'	1	1/2008
תכונות מנוע החיפוש Fast	מימוני אלון	טו'	2	6/2003
ארכיטקטורה של מנוע החיפוש Fast Search	מימוני אלון	טו'	2	6/2003
כריית מידע	פרופ' מימון עודד	יח'	1	7/2011
המוצר Autonomy	מנור עמית	כ'	1	6/2013
השוואה בין מימושים שונים של מורפולוגיה עברית ביישומי אחזור מידע טקסטואלי	מרגלית אפרים	י'	2	6/2004
NanoSyntax - גישה חדשנית להבנת שפה טבעית ביישומי מחשב	מרגליות ששון	יג'	2	6/2007

המשך אינדקס לפי שם מחבר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
הצגת תוצאות חיפוש מקובצת עפ"י טקסנומיה ארגונית בשילוב ממשק LCC&K	מרדכי ויקי	יג'	2	6/2007
הצגת תוצאות מקובצת של תוצאות חיפוש עפ"י טקסנומיה ארגונית בהתבסס על ממשק LCC&K	מרדכי ויקי דרורי עפר אריאל פרנק	יד'	1	1/2008
תקציר מעבודת מחקר בנושא "חברי הכנסת כצרכני מידע"	דר' מרקוס רבקה	יח'	2	1/2012
מאגרי מידע משולבים טקסט ומידע חזותי	סגל בני	ב'	2	11/1996
המדריך למנועי חיפוש ברשת האינטרנט	סימסולו יניב	ה'	1	1/1999
"חיפוש עברי: לראשונה בקוד פתוח. אתגרים, פתרונות, והתמודדויות אחרות"	סין-הרשקו איתמר	יז'	1	1/2011
עידון תהליכי חיפוש במאגרי מידע והצגתם למשתמש	סלייטר זיו חי-עזרא רחל	י"	2	6/2004
סמנטיק ווב וארגון ה-W3C	עידן אורי	טו'	1	1/2009
הרשת עושה היסטוריה: מוזיאונים משקיעים מיליונים במעבר לפורמט דיגיטלי	עילם הראל	יט'	1	6/2012
TDNet searcher analyzer	עפרון משה	יב'	1	1/2006
זיהוי טקסט וחיפוש בכתבי יד עבריים הסטורים	ערמון שחר	יט'	1	6/2012
גרפים קונספטואליים (CGS) דרך מעניינת להבנה סמנטית	דר' פוקס גיל	כ'	2	10/2013
ניצול אופטימלי של מנועי חיפוש	פישל אריק	ח'	1	1/2002
אחזור מידע פרלמנטארי מקבצי וידאו	פישל אריק	יח'	1	7/2011
שימושים ב-IR באתר השאלות והתשובות מהגדולים בעולם	פינשטיין יובל	טו'	2	6/2009
כלים ללימוד מכונה (Machine Learning) בביג דטה	פינשטיין יובל	כ'	1	6/2013
מגמות בזיהוי כתב יד	פלבינסקי מאיר	יט'	1	6/2012
מחשוב ארכיונים	פלבינסקי מאיר	יט'	2	12/2012
מנוע חיפוש עברי במסדי נתונים מובנים	פלמון ערן	ה'	1	1/1999
אינטרנט	פריימן שלמה	ב'	1	5/1996
ההתפתחות המקבילה של מנועי חיפוש וספריות דיגיטליות	פרנק אריאל חנני אורי	ז'	2	6/2001
מנוע החיפוש של SQL	פרנקל אסף	יט'	2	12/2012
תכונות מנוע החיפוש Flair	פרנקל עפרה	ט'	2	6/2003
מגלה סרקזם-ההמצאה הכי שימושית שהומצאה אי פעם	צור אורן	יח'	1	7/2011
תזאורוס, הרעיון ושימושו במערכות אחזור טקסט	צור מיכל	א'	2	10/1995
מנוע חיפוש כתשתית לאוטומציה של תהליכים ידניים במאגרי טקסטואליים	קולקו מיקי	ה'	1	1/1999
תכונות מנוע החיפוש XRS	קולקו מיקי	ט'	1	1/2003
אחזור טקסט בשמות באמצעות PowerMatcher	קולקו מיקי	י"	1	1/2004
מנוע לזיהוי ישויות	קולקו מיקי	טז'	2	6/2010
מנוע לניתוח קשרים של חברת 2001	קולקו מיקי	יא'	2	6/2005
אחזור מידע ומורפולוגיה של השפה הערבית	קמיר דרור	י"	2	6/2004
מפת הדרכים של החיפוש הארגוני במיקרוסופט - Fast	קסל תמיר	טז'	1	1/2010
ניתוח	קרימולובסקי יובל	יג'	2	6/2007
מערכת מודולרית לכריית מידע מטקסט	רגב יזהר, מאיה גורודצקי, רון פלדמן	יב'	1	1/2006
A Modular Information Extraction System	רגב יזהר, מאיה גורודצקי, רון פלדמן	יב'	1	1/2006
מיפתוח אוטומטי מול מיפתוח ידני בסביבה משרדית: בחינה השוואתית	רוזנברג תמי	יג'	1	1/2007
למה כל כך קשה לכתוב בעברית	רוזן יונתן	ה'	2	5/1999
לוח המפתחות העברי	רוזן יונתן	ה'	2	5/1999

המשך אינדקס לפי שם מחבר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כר ך	גליון מס.	תאריך
ערכים מספריים לאותיות עבריות	רוזן יונתן	ה'	2	5/1999
TRS - מאחורי הקלעים - המרכיבים השונים של המערכת והיישומים האפשריים בה	רוזן פטר	א'	2	10/1995
חיפוש באמצעות קלסיפיקציה של מידע	רזניקוב אלעד	טו'	1	1/2009
טכנולוגיות ה-DTSearch	ריפתין אביב	ח'	2	6/2002
שימוש במטה-תגיות לשיפור וייעול הופעת אתרים במנועי חיפוש	רפפורט טל רשתי דודו	ח'	1	1/2002
ניהול ידע	רפפורט טל רשתי דודו	ח'	1	1/2002
עברית ברשת	רשתי דודו	ה'	2	5/1999
דגשים בנוגע לשילוב עברית במסמכי HTML	רשתי דודו ירחי איציק	ה'	2	5/1999
הפקת מידע מתמונות דיגיטליות של קטעי הגניזה וזיהוי צירופים בין הקטעים	דר' שויקה רוני	יט'	1	6/2012
האם עיבוד שפות טבעיות הוא מסובך - ומדוע?	פרופ' שויקה יעקב	יא'	2	6/2005
המורכבות בפיתוח מנועי חיפוש בעברית	פרופ' שויקה יעקב	יא'	2	6/2005
מסמטאות קהיר לאינטרנט המהיר, מחשוב כתבי היד וזיהוי צירופים בין הקטעים	פרופ' שויקה יעקב	יט'	1	6/2012
אחזור מידע ומנועי חיפוש	שוק אדם	ו'	1	1/2000
Contex - מוצר לניהול תוכן	שושני יניב	יב'	1	1/2006
אחזור מידע המבוסס על תוכן תמונות במסמך	שטרן יוני	ג'	2	7/1997
ניהול תוכן במשרד במקר המדינה - ספריה דיגיטלית	שמחוני אלה	י'	1	1/2004
חוכמת האינטרנט-חשים ומכניסים הגיון לאינטרנט	שמיר דן	יח'	1	7/2011
מנועי חיפוש ארגוניים	שמעוני עינת	טו'	1	1/2009
ישום GSA במאגר פסקי דין	שמעוני גיורא	יט'	2	12/2012
שילוב מודלים של Web 2.0 בפרויקטי ניהול ידע	שמעוני עינת	טז'	1	1/2010
רשמים משולחן עגול בנושא Web 2.0 לצורך ניהול ידע ארגוני	שמעוני עינת	טז'	1	1/2010
סינון מידע בטכניקות מתקדמות	שפירא ברכה	ב'	2	11/1996
כלי לחיפוש שיתופי באינטרנט - Antworld	שפירא ברכה	ז'	2	6/2001
שיטות להתאמה אישית (פרסונליזציה) של תוכן	שפירא ברכה	י'	1	1/2004
מנוע חיפוש שיתופי מבוסס מודל כלכלי	דר' שפירא ברכה	יג'	2	6/2007
טכנולוגיה סמנטית להבנת שפה	שקד שאול	כי'	2	10/2013
XML - המסלול המהיר לכלכלה החדשה	שרוטר גרט	ז'	2	6/2001
כיצד להוציא מ-Google את המיטב	שרון טלי	יג'	1	1/2007
כלים לחיפוש מתקדם ברשת האינטרנט	שרון יוחאי	ה'	1	1/1999
מנוע אחזור באתר מוזיאון המדע בירושלים	שרון יוחאי	יא'	1	1/2005
בדיקת תוכנה נתמכת מחשב	שריג עידא	ד'	2	5/1998
Information Retrieval Interaction	Ingwersen peter	ט'	2	6/2003
Information Retrieval	C.J. van RIJSBERGEN	ט'	2	6/2003
אחזור מידע, פרק 1 - מבוא, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יא'	2	6/2005
אחזור מידע, פרק 2 - ניתוח טקסט אוטומטי, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יא'	2	6/2005
אחזור מידע, פרק 3 - סיווג אוטומטי, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יב'	1	1/2006
אחזור מידע, פרק 4 - מבנה קבצים, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יב'	1	1/2006
אחזור מידע, פרק 6 - אחזור הסתברותי, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יב'	2	6/2006

המשך אינדקס לפי שם מחבר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כר ך	גליון מס.	תאריך
אחזור מידע, פרק 7 - הערכה, תרגום חופשי C.J. van RIJSBERGEN של	C.J. van RIJSBERGEN	יג'	1	1/2007
אחזור מידע, פרק 8 - סיכום, תרגום חופשי C.J. van RIJSBERGEN של	C.J. van RIJSBERGEN	יג'	1	1/2007

אינדקס לכרכים א' עד כ' (כולל חוברת 2) עפ"י כותרים

אינדקס לפי שם מאמר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
אבני בניין לניהול ידע	מטה גרופ	י'	1	1/2004
אחזור טקסט בשמות באמצעות PowerMatcher	מיקי קולקו	י'	1	1/2004
אחזור מידע המבוסס על תוכן תמונות במסמך	יוני שטרן אבי עזרא	ג'	2	7/1997
אחזור מידע ומורפולוגיה של השפה הערבית	דרור קמיר	י'	2	6/2004
אחזור מידע ומנועי חיפוש	אדם שוק	ו'	1	1/2000
אחזור מידע פרלמנטארי מקבצי וידאו	אריק פישל	יח'	1	7/2011
אחזור מידע, פרק 1 - מבוא, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יא'	2	6/2005
אחזור מידע, פרק 2 - ניתוח טקסט אוטומטי, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יא'	2	6/2005
אחזור מידע, פרק 3 - סיווג אוטומטי, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יב'	1	1/2006
אחזור מידע, פרק 4 - מבנה קבצים, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יב'	1	1/2006
אחזור מידע, פרק 5 - אסטרטגיות חיפוש, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יב'	2	6/2006
אחזור מידע, פרק 6 - אחזור הסתברותי, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יב'	2	6/2006
אחזור מידע, פרק 7 - הערכה, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יג'	1	1/2007
אחזור מידע, פרק 8 - סיכום, תרגום חופשי מהספר של C.J. van RIJSBERGEN	C.J. van RIJSBERGEN	יג'	1	1/2007
אחזור מסמכים משובשים	משה לבנה	ח'	1	1/2002
אחזור קבצי מוזיקה ואודיו באינטרנט	פני בר סמן טוב	טו'	1	1/2009
איך לצמצם את שטח האחסון בדואר ג'יימיל	עפר דרורי	יח'	1	7/2011
אינטרנט	שלמה פריימן	ב'	1	5/1996
איתור נושא מסמך בצורה אוטומטית תוך שימוש במילים נפוצות	עפר דרורי	י'	2	6/2004
ארגון תוצאות חיפוש ממנועי אחזור באינטרנט	ד"ר מרק לסט	ח'	2	6/2002
ארכיון הכתבות הדיגיטליות בידיעות אחרונות	יוסי ברדוש	י'	1	1/2004
ארכיון צה"ל ומערכת הבטחון	אורי אלגום	ב'	1	5/1996
ארכיטקטורה של מנוע החיפוש Fast Search	אלון מימוני	ט'	2	6/2003
בדיקת תוכנה נתמכת מחשב	עידא שריג	ד'	2	5/1998
בחינת מודל ללמידה דרך הייפרטקסט	שמיר אחיטוב	ב'	1	5/1996
בניית מאגרי-מידע גדולים	עפר דרורי	ה'	2	5/1999
גרפים קונספטואליים (CGS) דרך מעניינת להבנה סמנטית	דר' גיל פוקס	כ'	2	10/2013
דגשים בנוגע לשילוב עברית במסמכי HTML	דודו רשתי איציק ירחי	ה'	2	5/1999
דירוג רשימת תוצאות חיפוש - ממצאי מחקר	עפר דרורי	ב'	2	6/2006
האם עיבוד שפות טבעיות הוא מסובך - ומדוע?	יעקב שויקה	יא'	2	6/2005
הבנת שפה ביישומי טקסטים	ד"ר עידו דגן	יג'	2	6/2007
הגניזה הקהירית ממשיכה לספק כותרות	דר' יכין אפשטיין	טו'	1	6/2012
ההתפתחות המקבילה של מנועי חיפוש וספריות דיגיטליות	פרנק אריאל ד"ר אורי חנני	ז'	2	6/2001
הוספת מנוע אחזור לאתר אינטרנט	עפר דרורי	ו'	2	6/2000
הוצאה לאור אלקטרונית	אמיר דנברג	ג'	1	2/1997
החפשו של סנונית	אוריאל אבולוף	ה'	2	5/1999
הטכנולוגיה הישראלית שמאפשרת להרכיב את הפאזל היהודי	עופר אדרת	יט'	1	6/2012
הטכנולוגיה של אוליב ל- OCR ושימוש ב-Fast	ירדן ירדני	יח'	2	1/2012
לאחזור מסמכי בתי משפט היפרטקסט וקבלת החלטות	רן גוראון	ד'	1	2/1998

המשך אינדקס לפי שם מאמר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
המדריך למנועי חיפוש ברשת האינטרנט	יניב סימסולו	ה'	1	1/1999
המוצר אוטונומי	עמית מנור	כ'	1	6/2013
המורכבות בפיתוח מנועי חיפוש בעברית	יעקב שויקה	יא'	2	6/2005
המרת אתרים מעברית ויזואלית ללוגית	אורלי ליס	ט'	2	6/2003
המרת שאילתות בשפה טבעית לשאילתות SQL	נועה אופק עפרי דר	יא'	1	1/2005
הנגשת ארכיונים לציבור הרחב באמצעות דוגמאות מיד ושם	מיכאל ליבר	טו'	2	6/2009
הסבת ארכיון צה"ל ממחשב WANG למחשב HP תחת UNIX	בני דננברג	ד'	1	2/1998
הפקת מידע מתמונות דיגיטליות של קטעי הגניזה וזיהוי צירופים בין הקטעים	דר' רוני שויקה	יט'	1	6/2012
הצגת תוצאת חיפוש בממשק משתמש במערכות אחזור טקסט - סקירת ספרות	עפר דרורי	ו'	1	1/2000
הצגת תוצאות חיפוש במערכות אחזור מידע תוך שימוש באלמנטים שונים לייצוג המסמכים - מחקר שימוש	עפר דרורי	יא'	1	1/2005
הצגת תוצאות חיפוש מקובצת עפ"י טקסונומיה ארגונית בשילוב ממשק LCC&K	ויקי מרדכי	יג'	2	6/2007
הצגת תוצאות מקובצת של תוצאות חיפוש עפ"י טקסונומיה ארגונית בהתבסס על ממשק LCC&K	ויקי מרדכי עפר דרורי אריאל פרנק	יד'	1	1/2008
הרשת עושה היסטוריה: מוזיאונים משקיעים מיליונים במעבר לפורמט דיגיטלי	הראל עילם	יט'	1	6/2012
"הרתחת אוקיינוס האינטרנט": הדור הבא של רשתות חברתיות	ד"ר יאיר יהודה	טז'	2	6/2010
השוואה בין מימושים שונים של מורפולוגיה עברית ביישומי אחזור מידע טקסטואלי	אפרים מרגלית	י'	2	6/2004
השוואה של מנועי חיפוש בעברית	תומר גיל	ח'	2	6/2002
השוואת מנועי חיפוש ומנוע ATTIVIO	צחי יעקובוביץ	טו'	2	6/2009
השוואת מערכות חיפוש של מאגר ה-Medline ויזואליזציה של מידע בהתבסס על קשרים מובנים וניתוחם	אלינור בנחקן	ה'	2	5/1999
השוואת מידע של מידע ברשת האינטרנט	פבל חרדק	טז'	2	6/2010
השוואת מידע של מידע ברשת האינטרנט	זמיר אורן	ז'	2	6/2001
"זה נשמע סינית? כלים לניתוח טקסט בעברית ובשפות אחרות"	דוד ליטוב אנדרו גולדמן	יז'	1	1/2011
זיהוי טקסט וחיפוש בכתבי יד עבריים הסטורים	שחר ערמון	יט'	1	6/2012
זיהוי תמידי של מידע ברשת האינטרנט	בועז חן	ו'	1	1/2000
זכרון לטווח רחוק	שגיא כהן	יט'	2	12/2012
חבילת מוצרים לניתוח ואחזור שמות חברי הכנסת כצרכני מידע	דביר בן אהרון	ט'	2	6/2003
חברי הכנסת כצרכני מידע	דר' רבקה מרקוס	יט'	2	12/2012
חדשות מקבוצת הענין	עפר דרורי	א'	1	4/1995
		א'	2	10/1995
		ב'	1	5/1996
		ב'	2	11/1996
		ג'	1	2/1997
		ג'	2	7/1997
		ד'	1	2/1998
		ד'	2	5/1998
		ה'	1	1/1999
		ה'	2	5/1999
		ו'	1	1/2000
		ו'	2	6/2000
		ז'	1	1/2001

המשך אינדקס לפי שם מאמר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
		ז'	2	6/2001
		ח'	1	1/2002
		ח'	2	6/2002
		ח'	2	6/2002
		ט'	2	6/2003
		י'	1	1/2004
		י'	2	6/2004
		יא'	1	1/2005
		יא'	2	6/2005
		יב'	1	1/2006
		יב'	2	6/2006
		יג'	1	1/2007
		יג'	2	6/2007
		יד'	1	1/2008
		יד'	2	6/2008
		טו'	1	1/2009
		טו'	2	6/2009
		טז'	1	1/2010
		טז'	2	6/2010
		יז'	1	1/2011
		יח'	1	7/2011
		יח'	2	1/2011
		יט'	1	6/2012
		יט'	2	12/2012
		כ'	1	6/2013
		כ'	2	10/2013
חוכמת האינטרנט-חשים ומכניסים הגיון לאינטרנט	דן שמיר	יח'	1	7/2011
חיפוש ארגוני, סיכום מפגש שולחן עגול	ליזה בודוגין	יט'	2	12/2012
חיפוש באמצעות קלסיפיקציה של מידע	אלעד רוניקוב	טו'	1	1/2009
חיפוש חופשי בטקסטים סרוקים בעברית	גיורא שמעוני	יא'	1	1/2005
חיפוש טקסט מלא בעברית ובערבית - הבעיה והפתרון	ליאורה ירדני	י'	1	1/2004
חיפוש לפי משמעויות	ד"ר אברהם מידן	יד'	1	1/2008
חיפוש מידע ברשת	קובי אדלשטיין וטל רפפורט	ח'	1	1/2002
חיפוש מידע מבוסס ניתוח טקסט	אורי וינהבר	יב'	1	1/2006
"חיפוש עברי: לראשונה בקוד פתוח. אתגרים, פתרונות, והתמודדויות אחרות"	איתמר סין- הרשקו	יז'	1	1/2011
חיפוש תלוי הקשר	חנן כהן	ח'	1	1/2002
טיפול בנושא השמות ביד ושם	אלכס אברהם	ט'	2	6/2003
טכנולוגיה סמנטית להבנת שפה	שאול שקד	כ'	2	10/2013
טכנולוגיות ה-DTSearch	אביב ריפתין	ח'	2	6/2002
טכנולוגיית ה-Push	טליה גליקשטיין	ג'	2	7/1997
טכנולוגיית מיקרוסופט לאחזור מידע (ביג דטה)	מאור דוד	כ'	1	6/2013
יין ואינטרנט	מרק טויטו	ו'	2	6/2000
ייצור אוטומטי של תיזאורי ומילונים דו לשוניים	איריס ארד	י'	2	6/2004
יציבות מידע ברשת	ד"ר יהודית בר-אילן	ו'	2	6/2000
ישום מערכת איחזור מידע טקסטואלי בשע"ם	עפר דרורי	א'	1	4/1995
ישום GSA במאגר פסקי דין	גיורא שמעוני	יט'	2	12/2012
כיווני פיתוח באחזור טקסט TQL	גלעד וייזל	א'	1	4/1995
כיצד להוציא מ-Google את המיטב	טלי שרון	יג'	1	1/2007
כיצד לייצר ספר אלקטרוני עם תכונות חיפוש באמצעות Adobe Acrobat	עפר דרורי	יא'	1	1/2005
כלי חיפוש סטנדרטיים במערכות ומוצרים קיימים	מוטי חגיז	ב'	2	6/2006
כלי לחיפוש שיתופי באינטרנט - Antworld	שפירא ברכה	ז'	2	6/2001

המשך אינדקס לפי שם מאמר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
כלים לביצוע מחקר איטרטיבי	יעל אמיתי	ג'	1	2/1996
כלים לחיפוש מתקדם ברשת האינטרנט	יוחאי שרון	ה'	1	1/1999
כלים ללימוד מכונה (Machine Learning) בביג דטה	יובל פיינשטיין	כ'	1	6/2013
כריית מידע	פרופ' עודד מימון	יח'	1	7/2011
לוח המפתחות העברי	יונתן רוזן	ה'	2	5/1999
למה כל כך קשה לכתוב בעברית	יונתן רוזן	ה'	2	5/1999
מאגרי מידע משולבים טקסט ומידע חזותי	בני סגל	ב'	2	11/1996
מבוא לדיבור ממוחשב, תרגום מכונה וזיהוי דיבור	דורון מודן	יז'	1	1/2011
מבוא לזיהוי דיבור	עידן בן שלוש	כ'	2	10/2013
מבנה המילה והשתקפותו בניקוד ובתעתיק	פרופ' עוזי אורן	יג'	2	6/2007
מגלה סרקזם-ההמצאה הכי שימושית שהומצאה אי פעם	אורן צור	יח'	1	7/2011
מגמות בזיהוי כתב יד	מאיר פלבינסקי	יט'	1	6/2012
מגמות עתידיות בעולם איחזור המידע expetrex	ד"ר אורי חנני	א'	1	4/1995
מורפולוגיה למנועי חיפוש	גורא שמעוני	יד'	1	1/2008
מחקר בתיק דיגיטאלי בבתי המשפט, כיצד?	ירדן ירדני	יח'	1	7/2011
מחשוב ארכיונים	מאיר פלבינסקי	יט'	2	12/2012
מחשוב המתווה הלקסיקוגרפי של השפה העברית בת זמננו	שירה זיסמן בן-דור	יא'	2	6/2005
מטריקה, סגמנטציה ומבנה קואורדינטות במערכת אחזור טקסט	איריס ארד	יב'	1	1/2006
מיומנו של מחפש עצמאי - איך מחפשים ומוצאים	זאב גרינברג	טו'	1	1/2009
מיפתוח אוטומטי מול מיפתוח ידני בסביבה משרדית: בחינה השוואתית	תמי רוזנברג	יג'	1	1/2007
ממנוע חיפוש לשלטי דרכים	עוזי אורן	יג'	1	1/2007
ממשק חלונאי אחד לטקסטים בסביבות עבודה שונות	ישראל מבשב כוכבה טל	ז'	1	1/2001
ממשק שפה טבעית בעברית למסכי נתונים יחסיים	ישראל מבשב כוכבה טל	ז'	1	1/2001
מנוע אחזור באתר מוזיאון המדע בירושלים - התלבטויות ושיקולים בבחירה	יוחאי שרון	יא'	1	1/2005
מנוע האחזור Inter Text - גרסה חדשה	ראובן אבגי	ט'	1	1/2003
מנוע החיפוש RetrievalWare	נחמה אנדלמן וצבי קמר	ט'	2	6/2003
מנוע החיפוש של SQL	אסף פרנקל	יט'	2	12/2012
מנוע חיפוש בטקסטים עבריים	עוזי אורן	יא'	2	6/2005
מנוע חיפוש וניהול ידע בעברית	עידית אופיר יורם זהבי	ח'	2	6/2002
מנוע חיפוש כתשתית לאוטומציה של תהליכים ידניים במאגרים טקסטואליים	מיקי קולקו	ה'	1	1/1999
מנוע חיפוש לשפה העברית	עפר דרורי	ט'	2	6/2003
מנוע חיפוש עברי במסדי נתונים מובנים	ערן פלמון	ה'	1	1/1999
מנוע חיפוש שיתופי מבוסס מודל כלכלי	ד"ר ברכה שפירא	יג'	2	6/2007
מנוע לניתוח קשרים של חברת 2001	מיקי קולקו	יא'	2	6/2005
מנועי חיפוש בעברית - רשימת ספקים	עפר דרורי	ח'	2	6/2002
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 3.2004)	עפר דרורי	י'	2	6/2004
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 4.2005)	עפר דרורי	יא'	2	6/2005
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 06.2006)	עפר דרורי	יג'	1	1/2007
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 7.2009)	דרורי עפר	טז'	1	1/2010
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 10.2010)	דרורי עפר	יז'	1	1/2011

המשך אינדקס לפי שם מאמר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
מנועי חיפוש ארגוניים	עינת שמעוני	טו'	1	1/2009
מנועי חיפוש באינטרנט	עפר דרורי	ג'	1	2/1997
מנוע לזיהוי ישויות	מיקי קולקו	טז'	2	6/2010
מנתונים לידע - MindCite	ד"ר אורי חנני	ט'	2	6/2003
מסמטאות קהיר לאינטרנט המהיר, מחשוב כתבי היד של גניזת קהיר	פרופ' יעקב שויקה	יט'	1	6/2012
מערכת לאיתור ישויות	אביבה יפת עפר דרורי	י'	1	1/2004
מערכת מודולרית לכריית מידע מטקסט	יזהר רגב מאיה גורודצקי רון פלדמן	יב'	1	1/2006
מערכת נוהלים בטכנולוגיית אינטרנט-נט	איציק הוד	ד'	2	5/1998
מפת הדרכים של החיפוש הארגוני במיקרוסופט - MOSS	רונה לוסטיג	טז'	1	1/2010
מפת הדרכים של החיפוש הארגוני במיקרוסופט - Fast	תמיר קסל	טז'	1	1/2010
ניאוזאורוס - מנוע חיפוש וניהול מימדים	עוזי אורן	יב'	1	1/2006
ניהול וארגון קבוצת ענין	עפר דרורי	ב'	1	5/1996
ניהול ידע	טל רפפורט דודו רשתי	ח'	1	1/2002
ניהול ידע בחקירות	נמרוד בלומקין	יח'	2	1/2012
ניהול ידע מחקר הצלי"שים באתר הגבורה	עפר דרורי	כ'	1	6/2013
ניהול תוכן במשרד מבקר המדינה	אלה שמחוני	י'	1	1/2004
ניהול תוכן - תכונות נדרשות	עפר דרורי	ט'	1	1/2003
ניצול אופטימלי של מנועי חיפוש	אריק פישל	ח'	1	1/2002
ניתוח שאילתות באתרי מכירות	ד"ר איגריד ולן	יג'	2	6/2007
ניתוח תחבירי חלקי	יובל קרימולובסקי	יג'	2	6/2007
סיווג אוטומטי של מסמכי טקסט בשפות שונות (כולל ערבית)	ד"ר מרק לסט	יג'	1	1/2007
סיכום מצב קיים במערכות אחזור טקסט	ד"ר רות הנדזל	ב'	2	11/1996
סינון מידע בטכניקות מתקדמות	ברכה שפירא	ב'	2	11/1996
סמנטיק ווב וארגון ה- W3C	אורי עידן	טו'	1	1/2009
עברית ברשת	דודו רשתי	ה'	2	5/1999
עיבוד שפה טבעית בערבית	נמרוד הרשברג כפיר בר	יד'	1	1/2008
עידון תהליכי חיפוש במאגרי מידע והצגתם למשתמש	זיו סלייטר רחל חי-עזרא	י'	2	6/2004
שם המאמר	שם המחבר	כרך	גליון	תאריך
עיצוב ממשק משתמש במערכות מידע	עפר דרורי	ה'	1	1/1999
עיצוב ממשק משתמש ל- WEB	צביקה ווידנפלד	ד'	2	5/1998
ערכים מספריים לאותיות עבריות	יונתן רוזן	ה'	2	5/1999
פיתוח מערכת טקסטואלית תומכת החלטה בסביבה מרובת פלטפורמות	אהוד אריאלי	ז'	1	1/2001
פרמטרים המסייעים להצלחת דיבור מוחשב	דורון מודן	כ'	2	10/2013
פתרונות לאבטחת איכות תוכנה	זוהר גילעד	ד'	2	5/1998
קיבוץ שאילתות במנועי חיפוש	גדי גולדרינג ואיתן פאר	י'	1	1/2004
קריטריונים להשוואה בין מנועי חיפוש	עפר דרורי	ז'	1	1/2001
קריטריונים לבחירת מנוע אחזור טקסט - גרסה 2	עפר דרורי	ט'	1	1/2003
קריטריונים לבחירת מנוע אחזור טקסט - גרסה 3	עפר דרורי	י'	2	6/2004
קריטריונים לבחירת מנוע אחזור טקסט - גרסה 4	עפר דרורי	יב'	1	1/2006
קריטריונים להשוואת מנוע חיפוש גרסה 5 - מאי 2009	עפר דרורי	טז'	1	1/2010
רשימת ספקים למנועי אחזור בעברית (3.2003)	עפר דרורי	ט'	2	6/2003
רשימת ספקים של מנועי אחזור בעברית (11.2002)	עפר דרורי	ט'	1	1/2003

המשך אינדקס לפי שם מאמר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
רשמים משולחן עגול בנושא Web 2.0 לצורך ניהול ידע ארגוני	עינת שמעוני	טז'	1	1/2010
שיטות להתאמה אישית (פרסונליזציה) של תוכן	ברכה שפירא	י'	1	1/2004
שילוב בסיסי נתונים ומאגרי - מידע ב- Web	עפר דרורי	ח'	1	1/2002
שילוב בסיסי נתונים ומאגרי מידע באתר ה- Web בספריה ובמרכזי מידע	עפר דרורי	ז'	2	6/2001
שילוב מודלים של Web 2.0 בפרויקטי ניהול ידע	עינת שמעוני	טז'	1	1/2010
שילוב מערכות אחזור טקסט ומערכות מידע קובנציונליות	עפר דרורי	ג'	2	7/1997
שימוש באונטולוגיות לניהול וארגון מידע	רינה ארד	יג'	1	1/2007
"שימוש בטכנולוגיות מבוססות קוד-פתוח ביישומי אחזור-טקסט	ליאונד גליבוב אורן יוסיפון	יז'	1	1/2011
שימוש במטה-תגיות לשיפור הופעת אתרים במנועי חיפוש	רפפורט טל דודו רשתי שולה גורן	ח'	1	1/2002
שימוש במילים נפוצות במסמך לאיתור נושא המסמך	עפר דרורי	ט'	1	1/2003
שימושים ב- IR באתר השאלות והתשובות מהגדולים בעולם	יובל פיינשטיין	טו'	2	6/2009
שיקולים בבחירת מערכת לניהול תוכן	עפר דרורי	יא'	2	6/2005
שרות אחזור טקסט של גוגל	עפר דרורי	טו'	1	1/2009
שרות החיפוש של גוגל	אמיר אליאסי	טו'	2	6/2009
תזאורוס, הרעיון ושימושי במערכות אחזור טקסט	מיכל צור	א'	2	10/1995
תכונות מוצר לניהול הידע D2K.NET	אייל כהן	ט'	1	1/2003
תכונות מנוע האחזור - ניאוזאורוס	עוזי אורן	יב'	1	1/2006
תכונות מנוע החיפוש מורפיקס	ליאורה ירדני	ט'	1	1/2003
תכונות מנוע החיפוש ATTIVIO	עפר דרורי	טז'	1	1/2010
תכונות מנוע החיפוש REXYGO	עפר דרורי	טז'	1	1/2010
תכונות מנוע החיפוש WIZ.DOC	אברהם מידן	ט'	1	1/2003
תכונות מנוע החיפוש XRS	מיקי קולקו	ט'	1	1/2003
תכונות מנוע החיפוש XRS	עפר דרורי	טז'	2	6/2010
תכונות מנוע החיפוש Fast	אלון מימוני	ט'	2	6/2003
תכונות מנוע החיפוש Fast	עפר דרורי	טז'	1	1/2010
תכונות מנוע החיפוש Flair	עפרה פרנקל	ט'	2	6/2003
תכונות מנוע החיפוש Inter Text	ראובן אבגי	ט'	1	1/2003
תקציר מעבודת מחקר בנושא "חברי הכנסת כצרכני מידע	ד"ר רבקה מרקוס	יח'	2	2012
Advisor - כלי ליצירת יישומי שפה טבעית	חברת סלנס	טו'	2	6/2009
A Modular Information Extraction System	יזהר רגב מאיה גורודצקי רון פלדמן	יב'	1	1/2006
Context - מוצר לניהול תוכן	יניב שושני	יב'	1	1/2006
Full Text - מנוע חיפוש בעברית	שחר אורן	ז'	1	1/2001
Guidance: כלי ליצירת ממשק גרפי למערכת MF	דרור אורנשטיין	א'	2	10/1995
Information Retrieval	Rijsbergen C.J. Van	ט'	2	6/2003
Informayion Retrieval Interaction	Peter Ingwersen	ט'	2	6/2003
Inter Text	איציק הוך	ד'	1	2/1998
MKnowledge - תוכנה לקבלה, ניתוח והצגה של ידע	ניר לוי	יז'	1	1/2011

המשך אינדקס לפי שם מאמר (כרך א' עד כרך כ' חוברת 2)

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
NanoSyntax - גישה חדשנית להבנת שפה טבעית ביישומי מחשב	ששון מרגליות	יג'	2	6/2007
RexyGo מנוע חיפוש חדש	יוסי דבש	טז'	1	1/2010
TDNet searcher analyzer	משה עפרון	יב'	1	1/2006
TRS - מאחורי הקלעים - המרכיבים השונים של המערכת והיישומים האפשריים בה	פטר רוזן	א'	2	10/1995
Web 3.0 מעבר לפינה - טכנולוגיות סמנטיות באינטרנט ובארגונים	רוני חזקיה	טז'	2	6/2010
WizDoc - מנוע חיפוש לפי משמעויות בעברית ובאנגלית	אברהם מידן	ט'	2	6/2003
XML - המסלול המהיר לכלכלה החדשה	שרוטר גרט	ז'	2	6/2001
XML והשלכותיו על בסיסי נתונים ואחזור טקסט	ראובן אבגי	ו'	2	6/2000
Yad Vashem names and places index	Alex Avraha	ט'	2	6/2003