



"מידע וטקסט"

עלון קבוצת עניין אחזור מידע וטקסט - SIGTRTS

1. חדשות מקבוצת העניין / עפר דרורי.....1

מפגשי הקבוצה:

2. כלי לחילוץ משמעויות מטקסט ע"י שמואל ובר ועוז של-בר.....5
3. רפורטואר, מערכת שאילתות מותאמת ללקוח ע"י גלעד ויזל.....11
4. חיפוש וחילוץ ישויות בסביבות שונות מבוסס
ניתוח מורפולוגי ע"י ליאור ארנן.....28

מאמרים:

1. סיכום מפגש שולחן עגול, ניתוח טקסט ע"י עינת שמעוני
2. סיכום מפגש שולחן עגול, אנליטיקה ע"י עינת שמעוני

תוכן עניינים

7. אינדקס לכרכים א' עד כא' (כולל חוברת 2) עפ"י מחברים.....59
8. אינדקס לכרכים א' עד כא' (כולל חוברת 2) עפ"י כותרים.....67

חדשות מקבוצת העניין

עפר דרורי



ט' בכסלו תשע"ה
1 בדצמבר, 2014

"מידע וטקסט"

עלון קבוצת עניין אחזור מידע וטקסט - SIGTRTS
חוברת 2 כרך כא' - דצמבר 2014

חדשות מקבוצת העניין

שלום לכולם!

אנו נפגשים שוב בעלון חדש נוסף. עלון זה פותח את כרך כא' של 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.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 לוויני (GPS) כונו אותנו לרחוב פועלי צדק 4,
ירושלים. נסיעה טובה!

מאז הוצאת החוברת האחרונה (כרך כא' חוברת מס. 1) ביוני 2014 נפגשה הקבוצה פעמיים.

המפגש הראשון התקיים ביולי 2014, במפגש נשמעו ההרצאות :

- א. כלי לחילוץ משמעויות מטקסט ע"י עוז של-בר
הצורך הוא לחילוץ משמעויות מתוך טקסטים רבים בשפות שונות בכלי אחד.
האתגרים שהוצגו :
- יצירת פלט אחד ללא תלות בשפת הקלט
- טיפול ברב משמעויות הנובעים מהקשרים שונים
כמו כן הוצגו דרכי התמודדות של הכלי אתם, כמו הבנה תרבותית של מבנה שמות ועוד.
- ב. רפורטואר, מערכת שאילתות מותאמת ללקוח ע"י גלעד ויזל
המטרה היא לתת לאנשים שאינם טכניים כלי ליצירת חיפושים מול בסיסי נתונים מסוג SQL לחיפוש מידע.
בפועל המערכת היא מתווכת בין SQL לבין הצרכים של המשתמש לאיתור מידע.

המפגש השני התקיים באוקטובר 2014, במפגש נשמעו ההרצאות הבאות :

- א. חיפוש וחילוץ ישויות בסביבות שונות מבוסס ניתוח מורפולוגי ע"י ליאור ארנן.
הוצגו הבעיות והמורכבות בשפה העברית. הוצג הפתרון של מלינגו לטיפול בשפה העברית במנועי החיפוש הארגוניים השונים.
הוצגו "פטנטים" להפגת רב המשמעויות למילים רבי משמעות כמו גם האתגרים המורפולוגיים שהעברית מעמידה בפני מפתחי מוצר חיפוש.
לצד מנגנוני החיפוש הוצגו גם דרכים לחילוץ ישויות מתוך מידע לא מובנה.
- ב. סקירת מוצרים לניתוח טקסט התומכים בעברית נכון לאוקטובר 2014 ע"י ערן הראל.
הוצג המצב הקיים היום ברשות איסור הלבנת הון ומימון טרור בכל הקשור למידע לא מובנה שנמצא ברשות. הוצגו הדרישות השונות של הרשות מכלי אחזור לקראת מכרז שהרשות יצאה אליו לצד התכונות הקיימות היום בכלים הקיימים.

3. מאמרים ועבודות

א. סיכום מפגש שולחן עגול, ניתוח טקסט ע"י ענת שמעוני

ב. סיכום מפגש שולחן עגול - אנליטיקה ע"י עינת שמעוני

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

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

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>

כלי לחילוץ משמעויות מטקסט

שמואל ובר ועוז של-בר

כלי אינטואיציה מלאכותית לחילוץ משמעויות מטקסט

שמואל ובר / עוז של-בר

אתגר א':

1. יצירת פלט אחיד, ללא תלות בשפת הקלט
2. פתירת רב משמעויות, הנובעות מהקשר שונה, או משלב ותחום שונה.

הפתרון של אינטו-ווי (פיתוח ה-IntuScan):

1. זיהוי השפה, המשלב, וכן זיהוי מילים המתועתקות משפה אחרת – ושחזור המילה המקורית.
2. קטלוג סטטיסטי לנושאים ותחומים הקשורים דתי וכו' – על מנת להפחית כפל משמעות.
3. בחינת הקשר, כלומר הטקסט שמופיע לפני ואחרי ביטוי מסוים – על מנת לפתור כפל משמעות.
4. קישור יחידות לקסיקליות לעצמים אונטולוגיים חד משמעיים, המשותפים ליחידות הלקסיקליות בכל שפות הקלט – כך ניתן ליצור פלט אחיד.
5. שימוש בעצמים שחולצו לביצוע קטלוג סטטיסטי נוסף (תתי קטגוריות).

אתגר ב': התאמת ישויות על פי שמן, על אף מגוון הוריאציות והשפות בהן השמות יכולים להופיע.

הפתרון של אינטו-ווי (Entity Matcher):

1. שימוש באלגוריתמים סטטיסטיים ובחוקים על מנת לזהות ישויות (בני אדם, ארגונים, מקומות וכו').
2. שחזור שמות מתועתקים לשמות בשפת המקור.
3. ניתוח השם וחילוץ מידע נוסף (מין, שם פרטי, משפחה, האב, שייכות שבטית ואתנית וכדומה).
4. אגרגציה של כל הוריאציות והמופעים של אותה ישות.
5. מציאת קשרים בין ישויות – כגון אב ובנו, אדם ומקום מוצאו וכו'.
6. יצירת מזהה לכל ישות מאוחדת, כולל כל הקשרים והמופעים.

Artificial Intuition Tool for Extraction of Meaning from Texts

Dr. Shmuel Bar, CEO IntuView

The Problem

Ever since the Tower of Babel, the human race has been occupied with translation to bridge the gap between languages, cultures, societies and nations. Translation serves many purposes: it enables us to broaden the scope of our cultural perspective, to see the world in a way that others – friends and foes – do, to retrieve ancient knowledge that, otherwise, would be lost to mankind and to communicate between people on a day to day basis. However, in a global environment challenged with enormous amounts of information, a challenge has arisen that cannot be solved by translation. This is the need to identify affinities and dis-affinities between semantic units in different languages in order to normalize streams of information and mine the “meaning” within them regardless of their original language. When we look for information or generate alerts – particularly in domains that are global – we should not be restricted to streams of information in one language; when we are interested in information – be it alerts on terrorism, fraud, cyber attacks or financial developments – we should not care if the origin is in English, French or Chinese. Methods to deal with this problem have generally been based on multi-lingual dictionaries that enable key words spotting (by input of a key word in one language, the search engine can add the nominal corresponding terms in other languages) or by automated translation of texts and application of the search criteria in the target language.

Current technology for automated translation leaves much to be desired. Existing automated translation technology and tools for document categorization can provide guidelines for translators or rough categorization of texts in different languages; none however can provide the user with real-time operable intelligence and a detailed understanding of the meaning of the document. The more esoteric the text is, the less automated translation and categorization can extract the sentiment or the “hermeneutics” of the text. Therefore current cross-language information extraction relies, in the end, on human translation and analysis of the content of each document before any operational decision. This process suffers from several major deficiencies: long processing times, high percentage of false negatives and false positives, loss of intelligence due to ignorance of central cultural nuances. Named entity recognition has also not achieved the holy grail of automated entity creation and relationary mapping of all entities, which may relate to the user’s requirement.

The need, therefore, is for technology that scans the entire gamut of information, identify the language and the language register of the texts, perform domain and topic categorization and match the information conveyed in different languages in order to create normalized data for assessment of the scope and nature of a problem. In the absence of a “silver bullet” of one technology that can be applied to all domains, the solution is based on emulation of the “intuitive” links that domain experts find between concatenations of lexical occurrences and appearances of a document and conclusions regarding the authorship, inner meaning and intent of the document. In essence, this approach looks at a document as a holistic entity and deduces from combinations of statements meanings, which may not be apparent from any one statement. These meanings constitute the “hermeneutics” of the text, which is manifest to the initiated (domain specialist or follower of the political stream that the document represents) but is a

closed book to the outsider. The crux of this concept is to extract not only the prima facie identification of a word or string of words in a text, but to expand the identification to include implicit context-dependent and culture-dependent information or "hermeneutics" of the text. Thus, a word or quote in a text may "mean" something that even contradicts the ostensible definition of that text.

A language is, in essence, a group of "dialects". The decision to call Swedish, Danish and Norwegian separate languages on one hand, and Moroccan, Libyan, Saudi Arabia and Egyptian all "Arabic" is political and not linguistic. Even within a language, the litmus test of inter-intelligibility is not always passed. For example: the correlation between the linguistic register of Shakespeare (or even a 19th Century writer) and the New York times may be no more than 60-70%. A modern English speaker would find it difficult to understand an "English speaker" from the Elizabethan era. This is also true regarding the register of a fatwa by al-Qaeda and a modern Kuwaiti newspaper.

Even within the same language register, words, quotations, idioms or historic references can be "polysemic"; they have different meanings according to the domain and the context of the surrounding text. **Waterloo** can be **a place in Belgium**, a reference to **a final defeat**, a **London railway station**, or a **song by ABBA**. It all depends in what context and who referred to it. A **verse in the Quran** may mean one thing to a moderate or mainstream Muslim and the exact opposite to a radical. The word "**court**" may have the following meanings: **a royal court**, a **sports court**, a **school courtyard**, a **courthouse**, or **to court a person**. In different languages, all these meanings may – or usually – are represented by different words altogether. When a person reads the word in a document, the word is immediately "understood" as holding the meaning typical of that type of document and the other meanings fade into the background. This process enables humans to "scan" texts without having to process each sentence to get a "gist" of the meaning and to identify the nature of the text. As the person accumulates more information through other features (statements, spelling, references) in the text, he either strengthens his confidence in the initial interpretation or changes it. A person who is accustomed to reading two newspapers – "The Telegraph", "The Times" and "The Guardian" – may be shown reprints without the name of the newspaper and recognize which newspaper it is from by the fonts and alignment. Even if he were shown the two articles retyped with the same font, he would most likely identify them by vocabulary, style and other features. Frequently, if asked, he would be hard put to explain his judgment.

Extra-linguistic knowledge also provides us implicit information on a name's bearer: gender, ethnicity, tribal and family relations, nicknames, status, religious affiliation and even age. We expect a person by the name of Nigel or Alistair, to be British and not American. We would also expect an American woman by the name of "Ethyl", "Lois" "Doris" or "Dorothy" to be an elderly woman. Indeed, the first two names are far more common in the UK than in the US and the female names were given to approximately 7% of the girls born in the United States in the 1930's and to less than 0.05% in the 1970's on.

IntuView has approached this problem through combining language-specific and language-register specific NLP (Natural Language Processing) with domain-specific ontologies. The IntuView technology¹ extracts such implicit

¹ See US patent - Decision-support expert system and methods for real-time exploitation of documents in non-english languages, US 8078551 B2, PCT/IL2006/001017

meaning from a text or the hermeneutics of the text. It employs the relationship between lexical instances in the text and ontology - graph of unique language-independent concepts and entities that defines the precise meaning and features of each element and maps the semantic relationship between them. As a result of these insights, the process of disambiguation of meaning in texts is based on a number of stages:

1. Identification of the “register” of the language. The register of the language may represent a certain period of the language), dialects, social strata etc. In the global world today, however, it is not enough to identify languages; the world is replete with "hybrid languages" (e.g. “Spanglish” written and spoken by Hispanics in the US; “Frarabe” written and spoken by people of Lebanese and North African origin in France and Belgium) that are created when a person inserts secondary language into a primary (host) language, transliterates according to his own literacy, accent etc. It is necessary, therefore, to take the non-host language tokens, discover their original language, back transliterate them and then find the ontological representation of that word and insert it back into the semantic map of the document.
2. Statistical categorization of a document as belonging to a certain domain, topic, or cultural or religious context in order to reduce ambiguity.
3. Use of the lexical tokens that directly precede or follow the token or tokens in question – in order to provide additional disambiguating information.
4. Based on the identification of the domain of the text, the lexical units (words, phrases etc.) are linked to ontological instances with a unique meaning (as opposed to words which may have different meanings in different contexts) that can be "ideas", "actions", "persons" "groups" etc. An idea may be composed of statements in different parts of the document, which come together to signify an ontological instance of that idea.
5. The ontological digest of the document then is matched with pre-processed statistical models to perform categorization.

This approach, therefore, is not merely “data mining” but “meaning mining”. The purpose is to extract meaning from the text and to create a normalized data set that allows us to compare the “meaning” extracted from a text in one language with that, which is extracted from another language.

This methodology is applied also to entity extraction. Here, the answer to Juliette’s question, “what’s in a name” is – quite a lot. A name can tell us gender, ethnicity, religion, social status, family relationships and even age or generation. In order to extract the information, however, we must first be able to resolve entities that do not look alike but may be the same entity (e.g. names of entities written in different scripts English, Arabic, Devanagari, Cyrillic) and to disambiguate entities that look the same but may not be (different transliterations of the same name in a non-Latin source language or culturally acceptable permutations of the same name). This entails: statistical modeling of the names to identify possible ethnicity in order to apply the appropriate naming conventions for disambiguation and matching; and then . extracting implicit and contextual information from that entity, matching and analysis and matching of names based on culture-specific naming conventions. This calls for identification of the likely ethnicity of the name (much the same way that humans intuitively understand that a name is Irish, Hispanic or Indian), back-transliterating it to the source language (if not written in its original language); validating the re-constructed names, parsing it for identification of constituent parts (given name, patronymics etc.) finding all its alternatives by application of cultural naming conventions and aggregating and matching the information

implicit in it or in its context in order to discover links between the entities.

One of the basic tasks of text analytics is named entity recognition - extraction from unstructured text of entities (persons, groups, locations, ideas and actions) which are relevant to the user's requirements, aggregation and matching of ostensibly separate entities, collection and validation of information on them, linking them to other entities of interest and adding them to the intelligence database for future use. The IntuView methodology automates this task while bridging the gap between languages. In essence, using algorithms based on knowledge of patterns and conventions of pertinent source languages and cultures, it recognizes entities within unstructured or structured texts, matches their occurrences in different sources and extracts information from the contexts in which they appear, creating a "virtual entity" which is uniquely identified by its definition as an instance within a domain-specific hierarchical ontology and by its internal attributes, which link it to other entities.

The tasks that the semantic analyzer performs include:

1. Identification of entities within a structured to unstructured text, using both statistical and rule-based algorithms to categorize them as possible persons, groups, locations, addresses (URLs, dates, bank accounts), ideas, actions, and basically every type which is defined in the ontology. The data base entity or unstructured text is analyzed, using NLP tools or data mining in the DB to determine if it contains relevant entities and what type of entities they represent.
2. Reverse transliteration of names back into the source language in order to determine the original form or forms of the entity. The identified named entity, written in any given script (Latin, Cyrillic, Hindi or Arabic, etc.) is analyzed to determine its possible linguistic origin and then processed to extract possible spelling variants in the language of origin of the name. Thus, the name of the entity is restored to the original name in the source language or possible source languages. Information from the transliteration, which provides further identification, is retained for later analysis.
3. Parsing of the named entity and performs cultural-linguistic sensitive analysis of its components. The analysis of these components provides further validation or reassessment of the categorization of the type of the entity (an entity which may initially be considered a person-entity may actually be a place entity named after that person). This engine fills the attribute slots of the virtual entity (gender, location, size, object, predicate, ethnicity etc.) to provide further qualities for identification and matching. The name is vetted to determine its validity (a possible name or corrupt one).
4. Aggregation of all the variants and aliases referring to an entity within the different inputs, while maintaining their source identity for possible regression.
5. Matching of entities by finding relations between identified entities (family relations in person entities, person-location relationships etc.).
6. Creation of a virtual "identity card" of an entity with all the aggregated information that is collected in the inputs about it.

Ultimately, we should be able to differentiate between situations in which we need to "translate" foreign language documents and those in which we only need to extract information from them. Translation is a tool – not an end unto itself - that should facilitate the ultimate goal of identifying relevant intelligence and sending it to where it may bring the most value. We believe that for a large part of the intelligence tasks, hermeneutic analysis, and automated

entity creation and matching and automated summarization will play a critical role in expediting the intelligence tasks.

Application of this concept to identification of threats in cyberspace would be based on the following:

1. Pre-generation of lexical based domain models for different supported languages and for various threat types based on previously identified texts. These models will be applied in crawlers on the traffic in the Internet in order to perform initial triage.
2. Analysis of the triaged texts to extract ontological digests.
3. Matching the ontological digests with models of corpora of texts that have been pre-categorized as reflecting different features.
4. Matching the ontological digests against each other to identify threads of communication (since the matching is performed on the ontological level it is cross-linguistic). These threads then are batched together.
5. Identification of a suspect entity in one document in the thread then can be the basis for unraveling the thread.

Application of the IntuView technology to identification of cyber-threats would greatly facilitate international cooperation of law enforcement and regulatory bodies against groups of international cyber-criminals.

רפורטואר, מערכת שאילות מותאמת ללקוח

גלעד ויזל

מערכת Reportoire - מערכת שאילתות מותאמת ללקוח

גלעד ויזל

תמצית

המערכת מאפשרת בנייה של מערכת אחזור המותאמת לישויות הארגוניות. המערכת מכוונת למשתמשי קצה המעוניינים לבצע שאילתות מורכבות שדורשות חיתוך נתונים ממספר ישויות ועל פי מספר סוגי קשרים אפשריים. המערכת מכילה רכיבים גנריים אך דורשת פיתוח מסוים בהגדרת הישויות.

המערכת מבצעת שליפות מבסיסי נתונים רלציוניים. כל ישות מוגדרת כ View ומבוססת על מספר טבלאות. הקשרים האפשריים בין הטבלאות המגדירות את ה View הינם קשרים פוטנציאליים ויתממשו בהתאם לשאילתא הקונקרטית.

בהגדרת המערכת מוגדרת גם רשימה של כל הקשרים האפשריים בין הישויות. משתמש הקצה יכול לבחור את שם הקשר הנדרש מתוך רשימה.

מהי הבעיה שאנו מנסים לפתור

מערכות מבוססות SQL מאפשרות ביצוע של שאילתות מורכבות. משתמשי קצה, לעיתים, נדרשים לבצע שאילתות מול המאגרים ללא עזרת אנשי IT. משתמשי קצה מכירים את הישויות הארגוניות ואת הקשרים האפשריים ביניהם. אולם, אין הם מכירים את המבנה הרלציוני. אין הם מכירים את שמות הטבלאות ומהם שדות הקשר. יותר מכך, אין הם מכירים את שפת SQL.

גם משתמשים המסוגלים לכתוב SQL יכולים לשפר את היצרניות שלהם. מאחר והרבה הגדרות חוזרות על עצמן וקשורות יותר למבנה הנתונים הארגוני ולא לשאילתא הקונקרטית של המשתמש. הגדרה ראשונית וחד פעמית של מערכת בסיס הנתונים הארגונית תקל על יצירת שאילתות.

ארכיטקטורת Reportoire.

המערכת מסופקת עם מערכת GUI מוכנה, מנגנון חילול SQL, ממשקים לסוגי בסיסי הנתונים השונים, הגדרות לשפות ומערכת הרשאות.

בשלב הגדרת המערכת מבוצעות הגדרת הישויות ברכיב הנקרא "מילון" או Dictionary. פעולה זו מתבצעת על ידי מומחה Reportoire לפי רשימת דרישות מהלקוח. כמו כן ניתנת תחזוקה שותפת הכוללת שנוי ההגדרות לאורך חיי המערכת.

מערכת הרשאות ממפה משתמשים לקבוצות. לכל קבוצה ניתן שם. במסגרת הגדרות הקבוצה אנו מגדירים את הערכים המסננים את הנתונים. דבר זה מאפשר הגבלת כל קבוצה לחלק מסוים של הנתונים. בנוסף לסינון הנתונים בשליפה, בקבוצת ההרשאות ישנה הגדרה של שדות האסורים בצפייה ובשליפה.

מערכת ה GUI מאפשרת הצגה של שאילתות קיימות. עריכה של שאילתות קיימות או הוספה של שאילתות. עריכה של שאילתא כוללת בחירה של ישויות, בחירת קשרים בין ישויות, ציון ערכים לברירת הנתונים ברמת שדה, בחירת סרגלי התצוגה הנדרשים וסדר המיון הרצוי. מערכת ה GUI מאפשרת תצוגה של מספר "דפים" כל אחד עם שאילתא אחרת. בנוסף לשאילתות "ראשיות" ניתן להגדיר ולבצע שאילתות תלויות הקשר. הכוונה היא ביצוע של שאילתא על סמך שורת נתונים בדוח. כאשר מסמנים שורה ניתן לפתוח רשימה של שאילתות אפשריות ולבצע אחת מהן. בצורה

זו אפשר "לטייל" בנתונים, כאשר תוצאות השאילתא המשנית יכולות להיות מוצגות באותו דף או בדף חדש.

ממשק לבסיס הנתונים זמין עבור Oracle ו SQL Server.

הגדרת View

View מוגדר מסדרה של טבלאות. אין להסיק מכך כי כל הטבלאות ב View משתתפות בכל שליפה. טבלה מסוימת יכולה להופיע במספר Views שונים. הטבלאות ב View מקושרות דרך שדות משותפים. קשר כזה יכול להיות מוגדר משדה בודד בטבלה אחת לשדה בודד בטבלה אחרת או מספר שדות בטבלה יכולים ליצור קשר אחד עם סדרת שדות מקבילה בטבלה אחרת. קשרים אלו נקראים "קשרים רופפים" מאחר ועדיין לא גובשו לכלל קשר ממשי.

טבלה "פיסית" אחת בבסיס הנתונים יכולה להופיע ב View מספר פעמים. למשל טבלת פענוחים המפענחת שדות שונים בטבלאות שונות.

בהגדרת ה View אנו מגדירים מהם השדות שיוחצנו כשדות שליפה. כמו כן אנו מגדירים מהם הסרגלים הזמינים באותו View ומהם השדות המרכיבים כל סרגל. בכל סרגל אנו מגדירים גם את שדות המיון.

בכל הגדרה של View נגדיר מהם השדות שיוחצנו כזמינים ליצירת קשר עם View אחר.

רשימת הקשרים החיצוניים ל Views מגדירה את הקשרים בין ה Views על פי אותם שדות קשר שהוחצנו בהגדרת ה View.

ה View עצמו הוא במבנה רשתי בו הטבלאות הם ה Nodes המקושרים דרך ה"קשרים הרופפים".

Query Composition

בתהליך הגדרת השאילתא המשתמש בוחר View אחר View ומגדיר את סוג הקשר ביניהם על ידי בחירת קשר מרשימת קשרים. קשר זה הוא "קשר ממשי" מאחר והמשתמש עצמו דורש אותו. לאחר הגדרת הקשרים הללו נוצרת רשת רחבה הכוללת מספר Views המקושרים על ידי קשרים ממשיים.

המשתמש מזין, בנוסף לכך, ערכים לשדות שליפה ובוחר סרגלים שיבנו את הדוח.

לפני בצוע השאילתא ובעצם לפני יצירת SQL מאותה רשת של הגדרות מתבצעת בחינה של הרשת המלאה. המטרה בתהליך זה היא לציין עבור כל טבלה ברשת האם היא נדרשת או לא. גם שדות בטבלאות עוברים תהליך בחינה שיסמן כל שדה כנדרש או כמיותר. המצאות קשר ממשי גוררת סימון שדות הקשר כנדרשים. המצאות שדות נדרשים גוררת סימון טבלה כנדרשת. קשר רופף יהפך לקשר ממשי במידה והתברר כי ניתוקו יגרור ניתוק של שני אזורים נדרשים ברשת.

שדות מסומנים כנדרשים במידה ומתבצעת שליפה עליהם או נדרשו כחלק מסרגל תצוגה.

תהליך זה מביא ל"גיוזום" חלקים לא נדרשים מהרשת המלאה. שיטה זו מביאה לכך שלא יתבצע קישור לטבלה שאין בה צורך באותה שאילתא קונקרטית. לפעולה זו יש משמעויות פונקציונאליות ולא רק שיפור בצועים.

בצועים

המערכת יכולה להיות מוגדרת על בסיס הנתונים עצמו או גזירה כלשהיא . ההחלטה על מה יתבצעו השליפות היא באחריות הארגון.

הבצועים במערכת Reportoire טובים מאחר ושלב הגדרת המערכת מאפשר את הגבלת אופן השימוש ונותן לארגון את השהות לבנות את מערכת האינדקסים הנדרשת לפי ההגדרות של מילון הנתונים.

שדות המיועדים לשליפה מוגדרים מפורשות בשלב ההגדרה וגם להם ניתן להגדיר אינדקסים כנדרש לפני הפעלת המערכת ולא לחכות להגדרות אינקרמנטליות בזמן הפעלת המערכת.

המערכת יוזמת יצירת SQL עם SubQuery במידה והדבר מתאפשר. בצוע של SubQuery יעיל מ Join חלופי.

סיכום

הפתרון המוצע על ידי Reportoire מביא לארגון פתרון שהותאם לישויות הארגון. משתמשי הקצה יכולים לבצע שאילתות מורכבות למרות שאינם אנשים טכנולוגיים.

ההגדרה הראשונית חוסכת הגדרות חוזרות ונשנות של משתמשים.

המערכת מגבילה את המשתמשים לשדות שליפה מסוימים וקשרים מסוימים. משתמש הקצה אינו יכול לחרוג מההגדרות הראשוניות. אולם מותרות בידו אפשרויות שילוב מגוונות ביצירת השאילתא הנדרשת.

המערכת מאפשרת יצירת מנגנון Data Discovery פשוט הניתן להגדרה על ידי משתמשי הקצה.

חיפוש וחילוץ ישויות בסביבות שונות מבוסס ניתוח מורפולוגי

ליאור ארנן וחיים גילהר

מלינגו: חיפוש וחילוץ ישויות בסביבות שונות

ליאור ארנן וחיים גילהר

בשל המבנה המיוחד שלה, השפה העברית מעמידה אתגר לא פשוט בפני מנוע חיפוש. השפה העברית מתאפיינת בריבוי רב משמעות, בכתיב מלא/חסר, ניקוד, מילים נרדפות וכו'. פעולת החיפוש במערכת ללא טיפול במאפיינים אלו תסבול מאחזור חסר או מאחזור תוצאות לא רלוונטיות.

הפתרון הוא לשלב במנוע החיפוש מנגנון חכם - רכיב מורפולוגי - ובכך לפתור את כל בעיות החיפוש בעברית ולשפר באופן משמעותי את תוצאות החיפוש.

הרכיב המורפולוגי הוא API המותאם למנועי החיפוש המובילים בשוק כגון Lucene\SolR, Attivio, Microsoft SPS & SQL, Idol, Oracle הסמנטית, חיפוש על פי סאונדקס (מצלול), טיפול בכתיבים שונים, חיפוש על פי מילים נרדפות (תזאורות), חיפוש ממוקד שמות, והוא מאפשר סימון מילות החיפוש בתוצאות.

הרכיב המורפולוגי פועל בצורה של נרמול – כל צורה מאלפי הצורות השונות תומר לערך יסוד אחד (צורת יסוד) הן בזמן אינדוקס החומר והן בזמן השאילתה, תוך ניתוח ההקשר אשר בה היא מופיעה, ובאופן כזה מתבצעת בסופו של דבר ההתאמה והאחזור.

הרכיב המורפולוגי מתממשק למנוע החיפוש באמצעות ממשק תוכנה (API) של מערכת האינדוקס והחיפוש של מנוע החיפוש כך שהוא מהווה חלק אינטגרלי ממנוע החיפוש. האינדוקס שייבנה יהיה אינדוקס מנורמל. בזמן השאילתה, מילת / מילות החיפוש מנותחות מורפולוגית (כמו בתהליך האינדוקס). תוצאת הניתוח היא ערך מנורמל המופנה לאינדוקס המנורמל. רמת הדיוק והרלוונטיות של האחזור מגיעה ל 97%.

תחום ניתוח הטקסטים אינו מסתיים בחיפוש. לקוחות רבים מעוניינים בכלי שיסייע להם להבין טוב יותר את הטקסטים שהם עובדים אתו, לעיתים גם במסגרת מנוע החיפוש. לצורך כך יצרה מלינגו API אשר ייתן מענה לניתוח מורפולוגי וחילוץ ישויות טקסטואליות. זהו ה- Intelligent Content Analysis

ה- ICA של מלינגו היא מערכת לזיהוי וחילוץ ישויות בטקסט בלתי מובנה בשפות עברית ערבית ופרסית.

התוכנה מחלצת מטקסט את הישויות המרכזיות המופיעות בו כגון שמות אנשים, מקומות, ארגונים, כתובות, מילים מקבוצות סמנטיות מסוימות, וכן מחרוזות חוץ לשוניות בעלות משמעות כגון מספרי טלפון, מספרי רישוי רכב, כרטיסי אשראי, כתובות דוא"ל, אתרי אינטרנט ועוד. בנוסף, מחזירה המערכת גם ניתוח מורפולוגי מלא ותלוי הקשר עם התגברות על רב המשמעות. הפלט מכיל מידע מלא על צורת בסיס, שורש, מספר, מין, ועוד. הישויות מחולצות לתקציר שבו הן ממוינות לפי קטגוריה. ה- ICA כולל אפשרויות להוספת והעשרת קטגוריות ארגוניות יכולות אלו מאפשרות שיוך מילים או שמות בשפה העברית לישויות (קטגוריות) חדשות או להוסיף לישויות קיימות.

דוגמה אחת לשילוב שני המוצרים ה- CS וה- ICA במערך החיפוש היא שימוש ב-facets, כאשר הפלאג המורפולוגי למנוע החיפוש מאפשר אחזור תוצאות מדויקות, ומערכת ה- ICA מאפשרת יצירה של סינונים עפ"י הישויות שנמצאו בטקסטים.

סיכום מפגש שולחן עגול, ניתוח טקסט

עינת שמעוני



סיכום מפגש שולחן-עגול

ניתוח טקסט – Text Analytics

מנחה

עינת שמעוני

לקוחות נכבדים שלום,

תודה על השתתפותכם במפגש שולחן עגול Round Table בנושא ניתוח טקסט – Text Analytics.

בתחילת המפגש, שמענו הרצאת אורח מרתקת מפי מר יואב פרידור, מנכ"ל באזילה. לא אסקור כאן את כל ההרצאה, אך אתן כמה highlights, שגם היוו קרקע לדיון בהמשך:

- באזילה פיתחה טכנולוגיה שמאפשרת לסרוק את המידע (שיחות ברשת), ובזכותה ניתן להתעסק ב-2 סוגי הזדמנויות:

1. LEADS - לשוחח עם אנשים ולגשת אליהם בזמן הנכון. הזדמנויות מסוג זה מביאות אוויר חדש

לעולם שלנו כמנהלי עסקים. "Lead" - השכבה הראשונה של המידע שהופך למכירות. למשל מישור שכותב בפורום משהו (שהופך אותו למעשה ל-lead) - אם נדע לתת לו ערך ספציפי רלוונטי, אנחנו מקרבים אותו למכירה. מנצלים את היכולת שקיימת כיום למצוא את השיחות האלו, ולתת ערך לאנשים.

2. להבין ולנתח את הסביבה שלנו. תחום זה מאוד מעניין ארגונים כיום. אפשר ללמוד הרבה מאוד

רק מלהקשיב: למה דברים קורים? מה אנשים רוצים? מהן התפישות שלהם?

- מדוע כל כך חשוב לנו לאתר לקוחות מתלוננים? הערך שנוצר מלפגוש לקוח מתלונן ולתת לו מענה טוב הנו ערך גדול מאוד - הוא מקצר מעגלי שירות, תורם למיצוב של העסק, למיקומו במנועי חיפוש ועוד. בפשטות, זה שווה כסף בזמן משבר, וחיסכון בעלויות בנזקים שנוצרו כתוצאה מהמשבר.
- על מנת להתמודד כיום עם המציאות העסקית המורכבת והדינמית, עולם המחקר עובר מעולם של מחקרים כמותיים, לעולם של מחקרי שאלה והקשבה.

מהם מחקרי הקשבה? זוהי קטגוריה חדשה בעולם המחקר, האם מדובר במחקר איכותני או כמותי? יואב הגדיר אותו כ"מחקר איכותני בכמויות": זהו אינו מחקר כמותי ואינו מחקר איכותני, אלא פשוט סוג חדש של מחקרים. אפשר להקביל זאת לצפייה בשבט באפריקה שמשוחח... אך עד כמה המחקר מייצג? יואב ציין כי מניסיונו, מחקרי הקשבה מייצגים את המציאות לחלוטין. רק במקרים חריגים יקרה שטוקבקיטיים "מטעים" ישפיעו על התוצאות.

- אחד היתרונות הבולטים במחקרי הקשבה הוא שמחקרים אלה עונים על שאלות שלא חשבנו עליהן אפילו, להבדיל מסקר רגיל עם שאלות מובנות. במחקר הקשבה אפשר פתאום לגלות כיוון מסוים, וללכת לחקור אותו. בהמשך הדיון שנערך לאחר ההרצאה, משתתפים התייחסו ליתרון זה כיתרון מרכזי ביותר, והדבר בהחלט תואם את המגמה הכללית בעולם BI/אנליטיקה - מעבר משאלות מובנות (Queries) לגילויים (Discoveries) ושיטות במידע. כשעושים סקר מחקר "סטנדרטי" ישנן שאלות שלא נחשוב לשאול אותן. לעומת מחקר הקשבה - מישור אמר משהו שנשמע מעניין, בואו נקוב אחרי זה...

מחקר הקשבה יכול לסייע במענה על השאלה החשובה ביותר כיום: מה אנשים תופשים? מה הם רוצים? מגמות?

- כמו כן, מחקר הקשבה יודע לנתח עבר – מה אנשים חשבו על המבצעים של חברה בפסח הקודם? אפשר להיכנס לאירוע מקביל בשוק מקביל, ולהבין איזה צעדים כדאי להם ליישם לקראת המהלך.
- בהרצאה ניתנה דוגמה למחקר שבוצע על קבוצת מחקר אימהות ב"גינה הוירטואלית" (פורומים בעיקר), תוך כדי המחקר התברר שישנם תתי-קהלים בתוך קבוצה זו, עם מאפיינים שונים (ולכן גם אסטרטגיית המכירה/פנייה אליהן צריכה להיות שונה):
 - אמא טרייה
 - אמא מסתגלת
 - אמא פז"מניקית
 - אמא ותיקה (פוזיציה מול הקהילה, מייצעות, קובעות סטנדרטים, עונות על שאלות לצעירות)
 - הקשבה לשיחות ב"גינה הוירטואלית" לימדה שיש תתי קהלים, שלא היינו יכולים להגדיר אותם קודם – מי שרוצה וחוזרת לעבודה, מי שחוזרת ולא רוצה, מי שלא חוזרת והייתה רוצה לחזור. כל אחד מאותם קהלים מתנהג אחרת. אלה דקויות נדרשות בעולם השיווק, כדי להביא מסרים יותר נכונים ולרכז את מאמצי המכירה בדברים הנכונים.
 - מחקר נוסף שבוצע על תחום שוק המזון בישראל – מחקר מגמות: מהם הכוחות שמניעים את המציאות העכשווית כדי לחזות מה יקרה בעתיד? לדוגמה, מגמת הטבעונות, אשר קצב ההתפתחות שלה כיום גדול מאוד. מחקר ההקשבה גילה כי כיום הטבעונות הנה יותר עניין של אכילה אתית, ופחות עניין בריאותי. טבעוני מתנסח בצורה של "אני אוכל את זה כי זה פחות הורג חיות", או "אני קונה את זה כי היחס לעובדים שם הוא טוב". גם אנשים שהם לא טבעונים לא מעוניינים לאכול בשר מחברה שמתייחסת לחיות בצורה מחפירה – לדוגמה ניתן להיזכר באירוע התאונה בו הייתה מעורבת משאית של חברת זוגלובק. היחס הניתן לחיות המובלות לשטיחה יצר צעקה כל כך גדולה, כזו שמשנה מבנה תפעולי של חברה, מצב תחרותי וכד'.
 - להבדיל מרוב המחקרים שהוצגו במצגת, בהן בהחלט מעורבת תשתית טכנולוגית אך לא פחות חשוב מכך – תהליכי מחקר, בהם יושבים חוקרים ומעבדים את המידע, הוצג גם מודל "אוטומטי" (בלחיצת כפתור) שנקרא BPI Brand Positioning Index – איפה המותג שלי ממוקם ביחס למחקרים אחרים? אפשר לעשות את זה ככלי – בלחיצת כפתור מקבלים דוח שלם. כלומר הוצגו גם מודלים אוטומטיים.
 - בהתייחסו לשאלות בנוגע לחסם העברית בישראל, ציין יואב כי אכן הנושא בעייתי (ניתוח סנטימנט בעברית מאוד בעייתי). בנוסף נשאלה שאלה - איך עוקפים ציניות והומור? דרך אחת היא- NLP ו Machine learning – אנחנו מלמדים את הכלי מה זה אומר כשמישהו אומר X... הכלי מייצר מודל מתמטי ומיישם אותו הלאה על תוכן וזה דורש עדכון כל הזמן. מבוסס חוקה. השיטות עוד לא הגיעו לשלמות. Machine learning אומר שאני אקח אלפי שיחות ואלמד את המחשב את החוקיות שבהן – מה חיובי, מה שלילי... המכונה תיישם את זה על שאר השיחות בעולם.

לאחר המצגת, בהמשך נערך דיון פתוח בין משתתפי המפגש – מקבלי החלטות שונים, רובם מתחום מערכות מידע וחלקם מתחום עסקי (שיווק), מארגונים גדולים בסקטורים שונים. ארגונים סיפרו היכן הם נמצאים כיום

בתחום יחסית חדש זה, מה התכניות שלהם בהמשך, התלבטויות וסוגיות, וחלקם – שכבר החלו צעדים ראשונים בתחום – חלקו מניסיונם וסיפקו כמה תובנות.

בין הסוגיות שעלו בדיון:

- האם כדאי ללכת בגישה של "לצרוך את נושא ניתוח הטקסט כשירות חיצוני" (במיוחד פיזיבילי כאשר מדובר על ניתוח מידע חיצוני – שיחות ברשת), או שהיתרונות בהחזקת נושא זה פנימית בארגון גוברים על מאמצי ההקמה והכניסה לתחום?
- האם הכלים תומכים בעברית בצורה מספקת? האם התחום בשל בישראל?
- ניתוח מידע בלתי מובנה פנים ארגוני (מסמכים / מיילים / תכתובות וכד') לעומת ניתוח מידע חיצוני (אתרים / פייסבוק וכד')?

מצ"ב סיכום עיקרי הדברים שעלו במהלך המפגש. במפגש עלו נושאים מהותיים שתומצתו בסיכום כפי שעלו. אין בסיכום זה המלצה גורפת ללקוחות אלא מתן פרספקטיבה והצגה של ההתלבטויות שעלו במפגש, כלומר "מהשטח".

קריאה מהנה,

בברכה, עינת שמעוני.

תוכן עניינים

- 6 מודל שירות חיצוני או הקמה פנימית בתוך הארגון?
- 7 כיצד להתניע פרויקט כזה?
- 7 האם התחום כבר בשל?
- 7 סוג המידע – איזה מידע ארגונים רוצים לנתח?
- 8 הכלים הטכנולוגיים
- 9 נספח – תגובות ספקים ויועצים לסיכום המפגש:

מודל שירות חיצוני או הקמה פנימית בתוך הארגון?

רוב הארגונים כיום צורכים את נושא ניתוח הטקסט ברשת (אתרים, פורומים, בלוגים, רשתות חברתיות) כשירות חיצוני, כיוזמה של מחלקת השיווק, שלא קשורה בכלל לאגף ה-IT. במפגש חלק מהמשתתפים סברו שזה אכן המודל המועדף ולא ראו ערך בהכנסת הנושא פנימה, בעוד שמשתתפים אחרים ציינו שהם רואים ביכולת ניתוח מידע לא מובנה יכולת אסטרטגית שיכולה להשפיע על היתרון התחרותי של הארגון בשנים הקרובות, ולכן יש יתרון להתחיל להתנסות, ובסופו של דבר כן לבנות את ה Practice הזה בתוך הבית.

- אחד הארגונים מתחום השירותים הפיננסיים תיאר 2 ערוצים:
 - ניתוח טקסט ברשתות חברתיות - הוחלט שנושא זה ינוהל כשירות אד-הוקי לאירועים ספציפיים (לא שירות שצורכים באופן קבוע).
 - איתור מידע פנימי – בתוך הארגון עצמו זורם המון מידע שחבוי בו ערך עסקי ממשי – עובדים שמתכתבים ביניהם, שיחות שנערכות בין לקוחות לנציגי הארגון (בין אם תמלול השיחה עצמה – voice to text, או תיעוד השיחה על ידי הנציג כטקסט חופשי ב CRM, התכתבויות בין עובדים ועוד). בתחום זה הכוונה היא ליישם תחום זה פנימית על ידי הארגון, ואף הותנע כבר פרויקט שמטפל בתחום זה.

ארגון נוסף מהתחום הממשלתי תיאר יוזמה של הזרמת מידע בלתי מובנה – ניתוח טקסט גלוי – והבאתו לתוך הארגון, תוך שילובו בפרויקט ניהול ידע שמתבצע בימים אלה. כלומר, בארגונים מסוימים נושא ניתוח הטקסט הנו בעיקר "משדרג" יכולות של מחקר מודיעיני קיימות. החוקר, בעבודתו המודיעינית, ייחנה מאינפורמציה נוספים שעד עכשיו לא יכל לקבל.

פרויקט שתואר - ניתוח עולם התוכן הקיים בארגון: מסמכים, ומתוכם לייצר תובנות. הכוונה היא להבנות מסמך לא מובנה, חילוץ ישויות וקשרים בין ישויות – אירוע, מקומות, ישות אדם / חברה, קשרים, פריטי מידע נוספים. התוצר - יוצא החוצה כמובנה אל המערכות שיודעות להתמודד עם זה כמובנה.

תוארה יוזמה נוספת על ידי ארגון פיננסי, גם היא רק בתחילת דרכה, לניתוח טקסטים שנרשמים בתוך ה CRM ולנסות להבין אירועי חיים/ נטישה... להבין סנטימנט.

כמה ארגונים ציינו רצון לתמלל שיחות call center אך לא היו מספיק מרוצים מהבשלות הטכנולוגית בישראל והתמיכה בעברית. הרצון לתמלל הוא לא רק כדי להפיק תובנות, אלא לעתים אפילו לדברים כמו – בדיקה האם הנציגים רשמו את מה שצריך ב-CRM.

כיצד להתניע פרויקט כזה?

אין ספק שתחום כזה, שהנו עדיין לא בשל כלל, ובהינתן נושא העברית הבעייתי, מצריך ביצוע בדיקת היתכנות / פיילוט, ואף כזה אשר אינו בהיקף קטן מדי (המלצה של אחד הארגונים שנכנסו לתחום). חשוב לבחון בפועל את יכולות הניתוח של הכלי, ובמיוחד לשים לב לנושא העברית. כל הכלים טוענים לתמיכה בעברית, ורובם אף מצינים שיש להם יכולת מורפולוגית (כל אחד בשיטתו הוא, חלקם על ידי התחברות למילון מורפולוגי קיים כמו זה של מלינגו, וחלקם על ידי פיתוח יכולת כזו משלהם או לחלופין גישת ה Machine Learning). אחד הארגונים מתחום הטלפון תיאר ניסיונות רבים לזהות אירועים והזדמנויות בזמן אמת – בזמן השיחה בין הנציג ללקוח, אך סיכם כי מאוד מורכב לדעת מה באמת קורה בזמן אמת ולהתמודד עם זה.

האם התחום כבר בשל?

התשובה היא "לא". האם זה אומר שכדאי לחכות כמה שנים? התשובה היא גם "לא". תחום זה כנראה הופך להיות אסטרטגי למיצוב התחרות של חברות. ארגונים שיחכו יהיו בפער משמעותי מדי מול ארגונים שמתחילים כבר עתה לצבור ניסיון.

אחד הארגונים הגדולים, בו קיים DW מתקדם מאוד, תיאר זאת בצורה הבאה: באבולוציה של ה-DW הקלאסי, הם הביאו אותו אל הקצה. המודלים הוותיקים של עולם ה-DW הקלאסי כבר הרבה פחות מוכיחים את עצמם מפעם. מחפשים משהו שיעזור להתמודד עם התחרות – שביעות רצון ומניעת נטישה. המודלים הקלאסיים שמבוססים על טבלאות רגילות – לא מספיקים. עולם הביג דאטה וניתוח טקסט הופך להיות יותר מעניין בעיני ה-BI והשיווק. ב-DW הקלאסי כבר נעשה הכל, חוץ מניתוח של VOICE וטקסט. כמו כן, נושא חשוב נוסף שיכול להוות קפיצת מדרגה אמיתית הנו ה- REAL TIME – לזהות התנהגות ושימושים ולהתאים בזמן אמתי ללקוח. כשעושים את זה נכון, זה מוכיח את עצמו.

סוג המידע – איזה מידע ארגונים רוצים לנתח?

ארגונים התייחסו לשני סוגים של מידע שברצונם לנתח:

- מידע "חיצוני" – השיחות ברשת. מידע זה משמש לבחינת מיקום המותג בסביבתו; זיהוי משברים (לדוגמה, חרם שהולך להתגבש) וטיפולם מבעוד מועד; זיהוי צרכים ורצונות - מה לקוחות רוצים, על מה מדברים; וגם באופן ספציפי – זיהוי לידים לצורך ספציפי (לדוגמה, מידע על "אירועים" בחיי לקוחות שיכולים להוות ליד – לדוגמה, מעוניין לרכוש דירה ולכן כדאי להציע לו ביטוח משכנתא בדיוור הקרוב). זיהוי לידים עלה כנושא חשוב, שכן ידוע שלקוחות מעוניינים לקבל הצעות שיווקיות רלוונטיות לצרכיהם ורצונותיהם, בזמן ובמקום המתאים. לצורך ניתוח מידע חיצוני, למעלה ממחצית מהארגונים שנכחו בדיון סברו שניתן וכדאי לנתח מידע זה באמצעות שירות של חברה חיצונית, ואין צורך להקים זאת בתוך הארגון.
- מידע "פנימי" – לדוגמה, המלל שנציגים מכניסים במערכת ה-CRM (שכיום אינו מנותח באף מקום); התכתבויות פנים ארגוניות / בין העובדים ללקוחות (עלתה כאן סוגיה אתית באשר לחוקיות שבניתוח מיילים של עובדים); תמלול שיחות הטלפון במוקד וניתוח תמלול זה (במפגש ציינו כמה ארגונים

שהנושא עדיין לא בשל מספיק בשל סוגיית העברית); מסמכים וחוזרים של הארגון ועוד. האווירה שעלתה מן הדיון היא שישנו ערך עסקי עצום במידע פנימי זה. רוב הארגונים יוזמים פרויקטים של הקמת נושא זה בתוך הארגון (ולא מסתכלים החוצה לשירות חיצוני).

אחת הנקודות שעוררה דיון (ונותרה ללא מענה מוחלט) היא – האם, מתי וכיצד ניתן (מבחינה משפטית) לנתח מיילים של עובדים? כמו כן, עד כמה הקניין הרוחני שהעובד מייצר שייך לארגון?

ארגון נוסף תיאר פרויקט של חיפוש ארגוני, במסגרתו טופל גם נושא תקנון ומדיניות חיפוש – אפשר יהיה להתייחס למידע של העובד אחרי שהוא עוזב כמידע ארגוני. לאחר שהעובד עוזב, מקפידים את המצב ולא נותנים לארגון לשנות, כך שאם מתעוררת בעיה אח"כ אפשר יהיה לחפש במידע זה.

הכלים הטכנולוגיים

אמנם מטרת הדיון לא הייתה לסקור את הכלים הטכנולוגיים, אולם עלו כמה שמות של טכנולוגיות וכלים שארגונים נתקלו בהם בחיפושיהם, חלקם כבר מיושמים בפועל וחלקם בדרך, נזכיר כאן כמה שמות שעלו (אולם, כאמור, אין זו רשימה ממצה כלל של הכלים):

- אטיביו Attivio
- HP – Autonomy
- IBM
- Melingo

נספח – תגובות ספקים ויועצים לסיכום המפגש:

תגובת חברת מטריקס:

שכיחות השימוש ב-text analytics מתגברת ככל שהטכנולוגיות הרלוונטיות משתפרות והופכות זמינות יותר. שימוש נפוץ הוא בתחומי חווית הלקוח לזיהוי מגמות צרכניות, סנטימנטים וכחלק מתוכניות Voice of the Customer. שימושים נפוצים נוספים הצוברים תאוצה הנם בתוך הארגון, כחלק מתהליכים של ניהול ידע ותוכן פנים ארגוני.

סיווג אוטומטי (המכונה auto categorization או auto classification) הוא פתרון העושה שימוש בכלי text analytics בכדי לתייג ולסווג פריטי מידע, מסמכים ותוכן. שימוש בסיווג אוטומטי בארגון משפר את היכולת להתמודד בצורה יעילה עם כמויות גדולות של מידע, לאכוף חוקים עסקיים ולשפר את חווית השימוש של משתמשי הארגון בבואם להשתמש במידע.

פריטי מידע ממקורות שונים, הכוללים מסמכים, מידע ממערכות ארגוניות, תכתובות דוא"ל ועוד, מסווגים כולם על בסיס סט מילונים משותף. חלק מהמילונים נבנים ומתוחזקים ע"י מומחי ידע בארגון ומאפשרים לזהות מילים ומושגים נפוצים אשר לכולם משמעות אחת ולאפשר סיווג משותף שלהם. לדוגמא: "מקולקל", "לא תקין" ו-"כשל" הן מילות מפתח שונות, המקושרות למילת המפתח "תקלה". כל מסמך המכיל אחת ממילות המפתח האלו (בהטיות לשוניות שונות) יקושר למילה "תקלה".

סט המילונים יכול גם פריטים המעודכנים אוטומטית ממערכות ארגוניות (כגון לקוחות, פרויקטים ועוד) ומרחיבים אותם כדי לגשר בין הזיהוי הרשמי של ישות המידע והאופן בו היא מכונה במסמכי הארגון. לדוגמא, פרויקט 05438 במערכת ה-ERP של הארגון, נקרא באופן רשמי "פרויקט החלפת מערכות ליבה" אבל בסלנג הארגוני מכונה גם "פרויקט תשתיות" או "הבלגן הגדול".

כמה תרחישי שימוש:

- פישוט תהליכי העבודה של המשתמשים בסביבת ניהול מסמכים – אין עוד צורך לתת מאפיינים למסמך.
- שליפה אוטומטית של מסמכים קשורים בתוך סביבת עבודה של פרויקט/מוצר/יחידה ארגונית.
- שימוש בחוקים עסקיים כדי לאכוף רגולציה על מסמכים המזוהים כרלוונטיים לנושאים מסוימים.
- ארגון אוטומטי של כמויות מסמכים גדולות במהלך הסבת מידע.

ועוד...

איש קשר : אדוה לוטן , ארכיטקטית יישומים ומנהלת מרכז התמחות ניהול ידע במטריקס

052-2385946 , edval@matrix.co.il

תגובת חברת IFN-יעל:

1. אכן קיימת דרישה הולכת וגוברת של לקוחותינו לבצע חיפוש טקסט בכלל מאגרי הארגון ולא רק במסמכים. לדוגמא, דוא"ל וצורפותיו, מסמכי SAP, תכתובת (chat) ממערכת ה-BPM. לכן, מערכת ה-ECM נדרשת ל: א. אירכוב כלל מאגרי התוכן ב. OCR לחומרים בעת העלאתם למערכת, לצורך חיפוש עתידי.
 2. בהחלט נדרשת לא רק יכולת חיפוש (מעין גוגל), אלא תובנה: ללמד את המערכת מהן תוצאות "טובות" לעומת "רעות" כך שניתן יהיה לקבל תשובה לשאלה: האם ראש העיר פופולארי? האם הציבור אוהב את הדגם החדש של יונדאי i35?
 3. עדיין קיימת בעיה בעברית. במייל שבו מופיעה המלה Sue, מנוע טוב ידע לזהות בקונטקסט באם מדובר בסוזן (הקיצור - Sue) או בתביעה – במקרה זה, בתוכן כזה יש לטפל מהר. מופע של "לחמו" בעברית יכלו להיות הלחם שלו או לחמו בהקשר המלחמתי...
- מנועי חיפוש של יבמ מופעלים כיום בהצלחה אצל מספר לקוחות IFN בישראל. במקביל, לקבוצת יעל יש פתרונות לשילוב חיפוש במסמכים במערכת IBM FileNet ממנועי גוגל ו-FAST.

איש קשר: משה (צ'יקו) יבניאל, משנה למנכ"ל, טל. 097638904, 0548016301

myavniel@ifn-solutions.com

תגובת חברת מלינגו:

באוטומציה של ניתוח טקסט, יש להבחין בין שתי רמות עיקריות

- a. איתור ופילוח: מציאת אזכורים רלוונטיים ללקוח בערימת השחת של מידע טקסטואלי, ופילוח אזכורים אלה לסוגים.
- b. מחקר: ניתוח כמותי של אזכורים – זיהוי מגמות, הפקת מסקנות מאזכורים מבלי שאדם יקרא אותם באופן פרטני.

איתור ופילוח

האיתור – הוא הגילוי, מתוך כמויות עצומות של אינפורמציה, של האזכורים והשיחות שראוי לארגון שלי להקשיב להן.

אם ניקח כדוגמה את עולם המידע ה"חיצוני" – כתבות, בלוגים, פורומים ורשתות חברתיות, ואת מלינגו כלקוח – מלינגו מעוניינת לגלות מקומות שבהם מוזכרים:

- מותגים: מורפיקס, רב-מילים, נקדן
- השווקים הרלוונטיים: מילונים, תרגום, מורפולוגיה, ניתוח טקסט
- מותגים מתחרים: בבילון, גוגל טרנסלייט
- אזכור אנשים: בכירים במלינגו, בכירים בתחום

איתור מפולח לנושאים:

מצאנו אזכור, למה הוא מתייחס? לבאג, לתכונה, למחיר, לדימוי החברה המוזכרת?

התוצאה מאיתור מוצלח ומפולח היא זאת:

מתוך זרם מידע של טרה-בייטים של חומר, נקבל את כל השיחות הרלוונטיות לארגון שלנו. אם הפילוח טוב, המידע יזרום לאדם הנכון בארגון (אם זה באג, ל-QA. אם תלונה על מחיר, למנהל המכירות, וכו'). יש לשים לב שבתום תהליך האיתור, הטיפול של החברה בכל מקרה הוא פרטני ואנקדוטלי. אין כאן מחקר כמותי, אין גרפים שמראים מגמות וכו'. יחד עם זאת, מטרת האיתור והפילוח היא מטרה ריאלית, אשר ניתנת לביצוע ברמת דיוק טובה מאוד בכלים העומדים לרשותנו היום.

נושא הטיפול היסודי במרכיב השפה הספציפית, לרוב עברית במקרה של השוק הישראלי (ובמידה מסוימת ערבית), הוא קריטי. יש צורך לענות על האתגרים הבסיסיים של השפה:

- הטייות (איפה אתם מתרגמים, לא הצלחתי לתרגם, למה התרגום לא טוב וכו')
- כתיבים שונים – טרנסלייט, טראנסלייט, טרנסליט
- מילים נרדפות ומילים שקולות – באג, תקלה, לא עובד לי, לא תקין (כל אחת עם הנטיית שלה)

יש לציין כדי לתת מענה ללקוח ספציפי לא מספיק כלי מרפולוגי out of the box, אלא יש צורך באדפטציה. הנרדפות של העברית הכללית אינן תמיד הנרדפות שרלוונטיות לארגון מסוים, וכמו כן יש לתת תמיכה לקסיקוגרפית באיסוף כל המילים שעשויות לציין תחום מסוים (למשל, בנק רוצה לאתר אירוע חתונה – אילו

מילים פרט לשורש "ח.ת.ב." יעידו על כך – זאת עבודה לקסיקוגרפית קלאסית ורצוי שתינתן כשירות ע"י חברת ניתוח הטקסט - מלינגו במקרה זה).

מחקר וניתוח כמותי

לעומת האיתור והפילוח, לגבי תחום המחקר הכמותי של מידע טקסטואלי עדיין קשה להגיד שהוא הגיע לבשלות.

כלומר, אם מלינגו רוצה לאתר את כל האזכורים של מורפיקס ולשלוח אותם ב-feed קבוע לאנשים הרלוונטיים, אשר יקראו אותם, ינתחו אותם וידווחו על בעיות, מגמות ומסקנות להנהלה – זה בהחלט יעבוד. לעומת זאת, אם מנכ"ל מלינגו רוצה לראות dashboard של אחוז שביעות הרצון מהמוצר מורפיקס לאורך זמן – גרף אשר ייוצר ע"י מחקר אוטומטי, ניתוח סנטימנט וכו' – זאת מטרה שהיא עדיין בעייתית נכון ל-2014.

יחד עם זאת, יש התקדמות גם בנושא זה. בדרך כלל, מדובר בשילוב של machine learning ושל NLP. כיום אילו אינן שיטות מתחרות אלא ברור למדיי ש-machine learning הוא חלק מ-NLP.

בגלל נושאי שפה שונים, בייחוד בשפה כמו עברית, גם כאשר משתמשים ב-machine learning, שכבה של ניתוח לשוני היא הכרחית. למשל, אם מנסים לאתר טקסטים שיכולים להעיד על נטישה, חשוב מאוד לנרמל צורות שונות של מילים רלוונטיות, שמביעות חוסר שביעות רצון. כל נירמול – השוואת צורות שונות שמדברות עם אותו הדבר, תומך בתהליך הסטטיסטי של ה-machine learning. יש גם שיטות שמבצעות שקלול של machine learning ושיטות rule based.

התחום הוא כאמור מאתגר אך אפשרי להגיע לתוצאות, כל עוד רמת הציפיות של הלקוח אינה מרקיעה שחקים, וכל עוד ברור שמדובר בחיסכון מסוים של זמן אדם אך לא במערכת שתקבל החלטות עבור הלקוח ללא מגע יד אדם. יש לקחת בחשבון שפרויקט בתחום זה הוא עתיר השקעה – ולתאם את הציפיות גם בנושא זה.

תגובת חברת SAP ישראל:

ל SAP יכולות מוכחות של Text analysis המשולבות בפלטפורמת ניהול המידע, SAP HANA . פתרון זה מאפשר ניתוח טקסט "On the fly" בביצועים מהירים בעשרות מונים מהפתרונות המסורתיים בתחום SAP. תומכת בניתוח טקסט ב 26 שפות ברמות ניתוח שונות, כאשר עברית היא אחת מהשפות הנתמכות. לקראת סוף השנה מתכננת SAP לשחרר תמיכה בעברית ברמה מתקדמת הכוללת יכולות מורפולוגיות מתקדמות המאפשרות ניתוח טקסט מתקדם לשימושים של ניתוח סנטימנט, שימושיים צבאיים, שימושי אחזקה ושירות לקוחות .

בפתרונות ניתוח הטקסט של SAP עושים שימוש לקוחות בעולם ובישראל במגוון מגזרים, החל מהמגזר הביטחוני, המגזר הציבורי והמגזר הפרטי.

פתרון ניתוח הטקסט של SAP תומך out of the box במיצוי ישויות ו Voice of the customer -בנוסף, בפתרון משולב מנוע חוקה המאפשר יצירת חוקים למיצוי ישויות נוספות (אשר לא מסופקות בפתרון) או שליפת עובדות (לדוגמה, בקשת לקוח, סנטימנט).

פלטפורמת SAP HANA משלימה את הפתרון באמצעות יכולות טיוב המידע והאינטגרציה, המאפשרים להביא את הישגיות שמוצו מהטקסט לרמת מידע איכותי שניתן לבסס ולקבל על בסיסו החלטות. כמו כן, הפלטפורמה מאפשרת לשלב את ניתוח הטקסט במנוע החיפוש של SAP HANA ובכך לאפשר חיפוש מתקדם בו ניתן לאתר קונספטים לדוגמה, כל המופעים של- החלפת מנכ"ל), זאת בנוסף לחיפוש טקסטואלי רגיל.

הפתרונות של SAP בנושא מאפשרים One stop shop עבור כל נושא הטיפול במידע הלא מובנה ובכך מאפשרים גישור על פערים ביכולות האינטגרציה הנדרשות מארגון ביישום פתרונות Text Analysis.

סיכום מפגש שולחן עגול, אנליטיקה

ענת שמעוני



סיכום מפגש שולחן-עגול

אנליטיקה – Analytics

מנחה

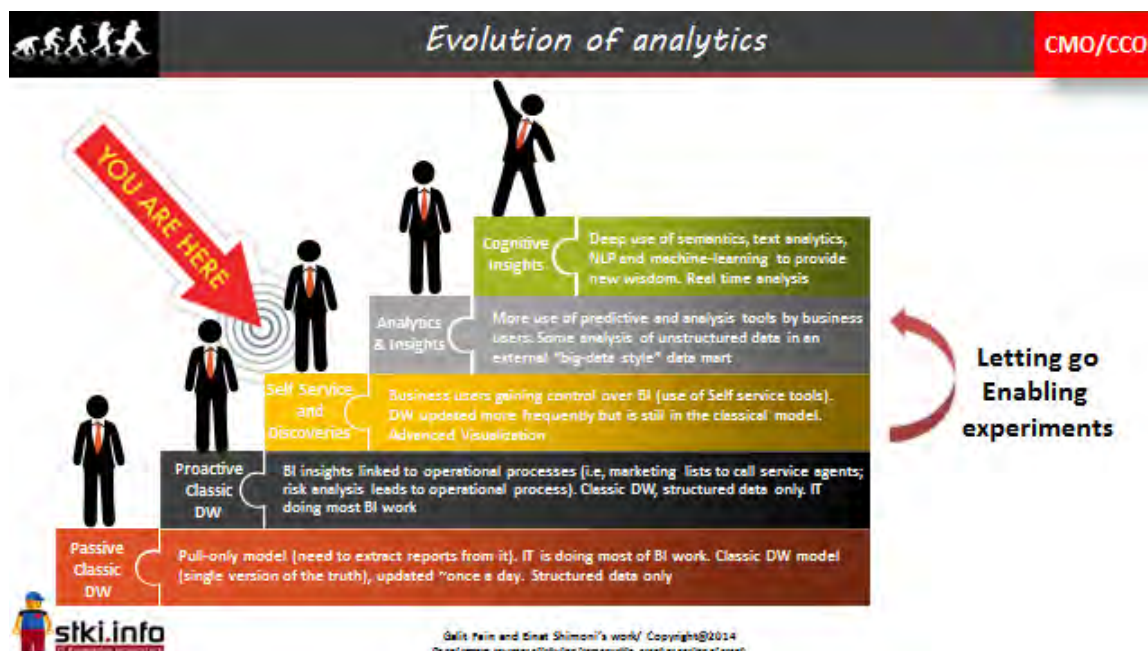
עינת שמעוני

לקוחות נכבדים שלום,

תודה על השתתפותכם במפגש שולחן עגול Round Table בנושא אנליטיקה - Analytics.

בתחילת המפגש הוצגו כמה שקפים על ידי STKI, הסוקרים את המגמות העיקריות בתחום. בין השאר, דובר על מודל הבשלות של תחום האנליטיקה, המתאר היכן רוב הארגונים בישראל נמצאים כיום, ולכן פניהם מועדות בתחום זה בשנים הקרובות.

באבולוציה של BI/אנליטיקה, רוב הארגונים מספרים על תחושת "רוויה" מתחום BI-DW המסורתי. מנהלי BI בשנה האחרונה התעסקו בשיפור יוזמות ה BI בכמה היבטים, בעיקר יישום יותר מודלים של Self service (מצב בו משתמשים מפתחים בעצמם דוחות ואף מודלים, לעומת מצב בו ה IT הוא מפתח הדוחות העיקרי), וכן מעבר לגישת ה-Discoveries (לעומת Queries). אולם הקפיצה העיקרית עליה מדברים מנהלי BI בשנים הקרובות היא הקפיצה לתחומים היותר אנליטיים, לספק יכולות תחקור יותר עמוקות, הכוללות אלמנטים של חיזוי, מגמות, חריגות, אופטימיזציות וכד'. כמו כן, יש לציין כי מדובר על אנליטיקה שונה מזו שהכרנו לפני כמה שנים במובן של הגבולות שלה. מדובר על ניתוח כל סוג של מידע, מובנה וכן בלתי מובנה, נפחים שאינם מוגבלים תיאורטית, ויכולות תחקור חדשות (לדוגמה, אנליטיקה "אוטומטית" שמספרת לנו מה קורה בנתונים שלנו).



בהמשך, בשולחן העגול נערך דיון פתוח בין משתתפי המפגש – רובם מנהלי BI בארגוני Enterprise, ומקבלי החלטות שונים, בסקטורים שונים. ארגונים סיפרו היכן הם נמצאים כיום בתחום זה, מהן התכניות שלהם בהמשך, התלבטויות וסוגיות, וחלקו תובנות על התחום.

בין הסוגיות שעלו בדיון:

- סוגיית ה"אמת המרכזית האחת" בארגון – כיצד מוודאים שתישמר, למרות כל ההתפתחויות הטכנולוגיות וסוגיית ה Self service?
- האם ארגונים בשלים ליישום מודלים סטטיסטיים, חיזוי וכד'?
- Big data עלה כתשתית מאפשרת בדיון (רוב הארגונים התייחסו אליו כאמצעי, ולא כמטרה).
- האם יש עדיין טעם ליישם "קוביות", או שהתשתיות האנליטיות החדשות, מבוססות in memory, מייתרות נושא ותיק זה?

מצ"ב סיכום עיקרי הדברים שעלו במהלך המפגש. במפגש עלו נושאים מהותיים שתומצתו בסיכום כפי שעלו. אין בסיכום זה המלצה גורפת ללקוחות אלא מתן פרספקטיבה והצגה של ההתלבטויות שעלו במפגש, כלומר "מהשטח".

קריאה מהנה,

בברכה, עינת שמעוני.

תוכן עניינים

5	הגיע הזמן לאנליטיקה!
5	משתמשים שונים – צרכים שונים
5	איך מתחילים? הגישה המסודרת אל מול הגישה הניסיונית
6	איום על "אמת אחת"
6	כלים טכנולוגיים
7	נספח – תגובות ספקים ויועצים לסיכום המפגש:

הגיע הזמן לאנליטיקה!

משתתפי המפגש בהחלט ציינו כי הם מרגישים שה-BI המסורתי וה-DW הקלאסי מתקרבים למיצוי. הם מרגישים שתוך שנה-שנתיים זה כבר לא יספיק והם יידרשו לתת תובנות מסוג חדש לארגון, כאלו הנשענות על כלים יותר אנליטיים, ניתוח סוגים חדשים של נתונים, וכן הכללת נושא החיזוי – Predictive analysis, כאשר ברקע – טכנולוגיות ביג דאטה מאפשרות הרחבת היריעה של האנליטיקה הקלאסית כפי שרובנו מכירים אותה, ומאפשרות ניתוח מסה גדולה של נתונים, מגוון רחב של סוגי נתונים ועוד. בין היוזמות שארגונים תיארו אשר הם רואים בהם ערך רב, והיו רוצים ליישם בארגונם:

- תמלול השיחות הטלפוניות במוקד השירות, ניתוח שיחות אלה להוצאת לידיים, זיהוי לקוחות בסיכון נטישה ועוד.
- ניתוח טקסט – חילוץ ישויות וקשרים בין ישויות (על גבי מסמכים, מיילים)
- עולמות אנליטיים "קלאסיים" כמו סגמנטציה, חיזוי נטישה, חיזוי הכנסות מול הוצאות, Collection – רמת ודאות להגעה אל הכסף, סיכוני אשראי, חישוב עושר פיננסי ללקוח

ארגונים סיפרו על יוזמות שונות שניסו להתניע בארגון בנושא האנליטיקה. בין היוזמות שתוארו - ניסיון לניתוח שיחות טלפוניות במוקד השירות – סופר על ניסיון שלא הצליח, בגלל חוסר בשלות התחום ותמיכה לא מספיק טובה בעברית.

משתמשים שונים – צרכים שונים

ברור שלא כל משתמשי הארגון ישתמשו ביכולות אנליטיות. קיים צורך למפות סוגי משתמשים שונים כאשר תהיה אולי סביבה ארגונית אחת לנתונים, אך כל אחד יעשה בה את השימוש שלו. לדוגמה, אחת מחברות הביטוח תיארו את האקטואריה ואנליסטים כמשתמשים פוטנציאליים של סביבות אנליטיות מבוססות in-memory (אותם אקטוארים משתמשים בכלי סטטיסטיקה וכריית נתונים כיום, אך באופן נקודתי ולוקאלי – על המחשב שלהם ולא בהיקף Enterprise).

בצד השני, יש ארגונים בהם רוב המשתמשים עדיין ב Mindset של דוחות ארוזים בקופסה. המשתמשים לא מבקשים בכלל Self-service ולא ברור מי (אם בכלל) ישתמש בכלים כאלה, אם הארגון יספק אותם.

איך מתחילים? הגישה המסודרת אל מול הגישה הניסיונית

רוב משתתפי המפגש, כאמור, ציינו כי ברצונם להיכנס לעולם האנליטי, חלקם גם ציינו מפורשות את תחום הביג דאטה כתחום שברור שצריך להתחיל לבדוק אותו כיום, אולם רובם ככולם מתקשים לאתר את "הפרויקט" שיצדיק כניסה לתחומים אלה. כלומר, אין כאן מצב בו מגיע משתמש עסקי עם צורך Business Case ברור, אלא המצב הוא שמנהלי BI מרגישים שיש כאן קפיצה טכנולוגית שיכולה לספק לארגון תועלות חדשות שהוא אינו מכיר, אך כיצד להתחיל להתניע מהלך כזה? לאיזה צורך עסקי כדאי "להתחבר"? כבר

ברור להרבה ארגונים שלא כדאי לשבת ולחכות שיגיע מנהל עסקי עם רעיון מגובש, אלא צריך להבהיר לארגון מה הוא יכול לעשות עם כלים אלה.

גישה אחרת, שאחד המשתתפים תיאר במפגש, היא הגישה הנסיונית. פשוט להתחיל "לשחק" עם התחום. באחד הארגונים תואר מצב בו מסת הנתונים הייתה גדולה מדי, ולכן התחילו להוריד Hadoop ולהתנסות, הקימו מערכות מעל, עשו שימוש בכלים / יוזמות קוד פתוח כדוגמת Couchbase ו MongoDB, הכשירו אנשים פנימית, וככל שנחשפו לזה בארגון, "עם האוכל בא התיאבון" והם התחילו לקבל דרישות. בשכבה האנליטית, חוקרי data scientists, מוכוונים משימות: מקבלים בעיה, מפתחים מודל, מכניסים בתוך אפליקציה וממשיכים הלאה. אותו נציג הביע אמון רב בכלי קוד פתוח ככלים העיקריים כיום בתחום זה, יותר מכלי ה Enterprise בצורה משמעותית. MAPR ו Cloudera מתקדמים יותר ויותר לכיוון סוויטה שלמה שעושה הכל – האדופ, אנליטיקה, ויזואליזציה... כיום קיימת סביבת Hadoop Sandbox, אשר חיה לצד ה-DW המסורתית.

איום על "אמת אחת"

הגישות האנליטיות החדשות, וגם סביבות ה-BI החדשות, מאיימות על הסטטוס-קוו של שמירת "Single Version of the Truth" שמנהלי BI כ"כ התאמצו להגיע אליו. גישת Self service מייצרת "איים" של מידע, ופרשנויות שונות של נתונים. אנליסטים / משתמשים שונים מייצרים בעצמם מודלים ולכן מגיעים לתוצאות עסקיות שונות. ניסיונות שארגונים תארו על מנת להתגבר על אתגר זה:

- בשכבה האנליטית, ארגון פיננסי בנה מנגנון לאנליסט שיושב במחלקה העסקית, שעובד מול מסד נתונים גדול של ה-DW שלהם – כשאותו אנליסט רוצה להפוך מודל חכם שהוא בנה, למשהו "תפעולי" שה-IT יארוז, אשר ירוץ כל הזמן, יש מנגנון שהוא עובר ומגיע ל IT (עוטפים אותו, כולל תמיכה בהרשאות).
- בשכבה של ה-BI, בתחום self service מגמה שהטמיעו די חזק – ב IT מייצרים שכבות סמנטיות, עוטפים אפליקציות ומנגישים את המידע למשתמשים החכמים / אנליסטים.
- ארגונים גם דיברו על הגדרת מילון נתונים כדרך להתמודדות עם הביזור הטבעי שמתרחש כעת בעולם ה-DW – הגדרת מילון נתונים על כל טבלה וכל שדה. מטפל גם בנושא המשמעות של נתון.
- יצירת "קהילת BI" בארגון שנפגשת מדי פעם, עושים כנסים והדרכות כיצד נכון להשתמש.

כלים טכנולוגיים

בין היתר, כלים טכנולוגיים מעניינים שהוזכרו (זוהי כמובן לא רשימת כלים מלאה), לא כולם מוכרים בישראל, וחלקם בקוד פתוח:

- כלי בשם [WEKA](#) – כלי קוד פתוח ל machine learning, בונה מודל חיזוי, classification, קלסטרינג, רשתות בייסיאניות
- לצד השימוש בכלי SAS, SPSS, הוזכר גם מעט שימוש בכלי R. הוזכר גם כלי קוד פתוח שנקרא PSPP (שלא במקרה מזכיר את SPSS)
- KNIME כלי ETL קוד פתוח בעולם הביג דאטה, גוררים ריבועים ומושכים חצים (ריבועים חכמים שעושים קלסיפיקציה וקלסטרינג)

נספח – תגובות ספקים ויועצים לסיכום המפגש:

תגובת חברת מטריקס:

SAP מספקת בשנים האחרונות פתרונות אנליטיקה מלאים המכסים כל צרכי האנליטיקה של הארגונים, החל מ Enterprise BI באמצעות סוויטת ה-BI של SAP Business Objects, פתרונות ה-SAP Lumira Self Service BI ופתרונות האנליטיקה המתקדמת SAP Predictive analysis.

פתרון SAP Lumira נועד לאפשר Self Service BI באופן שמאפשר יצירת "רשת של אמת", ע"י יצירת הניתוחים על בסיס עולמות והעשרתם במגוון נתונים רחב, החל מגיליונות נתונים, קבצים, בסיסי נתונים אחרים, אפליקציות ענן ושירותי רשת. הפתרון הינו פתרון קל לשימוש ואינטואיטיבי המאפשר זריזות ארגונית תוך כדי שמירה על בטחון במידה ועל השכבה הסמנטית הקיימת בתשתיות ה-IT. גישה חדשנית זו מהווה יתרון מהותי מול גישות ה-Self Service BI בשנים האחרונות.

SAP Predictive analysis הינו פתרון כריית מידע המבוסס על רכישת חברת KXEN בשנת 2013. פתרון זה הינו פתרון קל לשימוש, מהיר, בעל Time to market הקצר בעשרות אחוזים לעומת שימוש בכלי Opensource או כלי כריית המידע המסורתיים בשוק. הפתרון הינו פתרון אשר מאפשר לאנליסט פשוט שאינו בעל ידע סטטיסטי מקיף לבנות מודלי חיזוי, קליסיפיקציה, מנגנוני המלצה וכדומה במהירות ובקלות רבה ללא כל תלות במקור המידע. הפתרון מיושם אצל מאות לקוחות בעולם במגוון תעשיות ולמעשה מגשר על הפערים הקיימים בין רמת הידע הנדרשת עבור מדען מידע לעומת האנליסט העסקי שיכול להפעיל את הפתרון של SAP.

איש הקשר: בעז גודוביץ, מנהל מוצר sap business objects, boazg@matrix.co.il

תגובת חברת אמת מחשוב:

ארגונים רבים מוצאים ש-Splunk, מוצר לניתוח מידע IT, מאפשר מענה מהיר וסקאלאבילי לאתגרי אנליטיקה רבים. מלבד מערכת הפעלה, אין צורך בתשתיות תוכנה נוספות. המוצר כולל את כל הרכיבים הדרושים - ניהול מלא של מחזור החיים של מידע לאנליטיקה, מנוע תשאול וחיפוש, תשתית וויזואליזציה מתקדמת ולאלו המיישמים אנליטיקה בזמן אמת - מנגנוני תגובה לאירועים חריגים. מול מערכות מבוססות RDBMS, יישום Splunk לאנליטיקה מאפשר לדלג על שלבי ניתוח המידע ויישום ETL. מול יישומים מבוססי Hadoop, הטמעת Splunk מהירה לאין שיעור וכוללת, במוצר אחד, את כל הנדרש כדי לאפשר תחקור ע"י מדעני המידע בארגון. באמצעות Spunk ניתן לתחקר כמויות מידע עצומות באמצעות שפת חיפוש עשירה הכוללת יכולות סטטיסטיות נרחבות. מתוך שפת החיפוש ניתן במידת הצורך להפעיל תוכניות R או מודולי הרחבה שונים (כגון - הרחבות ל-Learning Machine, חיפושים צולבים במאגרי מידע מקומיים או מרוחקים ועוד). עם כניסת מקורות מידע בקצב גבוה כמו יישומי אינטרנט מבוססי ענן, יישומים להתקנים ניידים, ו-Internet of-Things, מערכות אנליטיות צריכות להתמודד עם שינוי מהיר במידע, בנפחים ובתובנות הנדרשות מהמידע. Splunk מאפשר ההתמודדות זריזה ואפקטיבית עם שינויים כאלו מעצם היותו מוצר דינאמי (ללא אכיפת סכימה, שליפת שדות בזמן חיפוש בלבד) ובעל יכולת גידול בנפח וביצועים באמצעות תוספת שרתים בלבד תוך כדי פעולה. תכונות אלו הופכות את Splunk למוצר אנליטיקה בעל מוכנות גבוהה מאוד לעתיד והגנה משמעותית על ההשקעה.

איש הקשר: אלי קפלן, elik@emet.co.il, 054-5771106

תגובת חברת IBM:

בעקבות שינויים מהותיים ודינמיים המתרחשים כיום בסביבה החיצונית והפנימית בארגונים, נדרשים מנמ"רים ומנהלי BI בארגונים לתת ערך מוסף עסקי-אנליטי ליחידות העסקיות השונות. למעשה, נוצר מתח בין חדשנות ויצירתיות מול תפעול ואופרציה של מערכות אנליטיות קיימות. מעניין לציין, לדוגמא, כי פתרונות מתקדמים, כגון פתרון ניהול הונאות (Counter Fraud Solution) המוצג בהמשך, משלבים כלי אנליטיקה מתקדמת מהעולם ה"חדש" ומהעולם ה"ישן", בצורה אינטגרטיבית - על-מנת לתת מענה ורטיקלי מתקדם בתחום. בנוסף לכך עולה השאלה הארכיטקטונית – כיצד לבחור כלים שבסופו של דבר יאפשרו:

- ניהול ארגוני בסטנדרטים הנדרשים (למשל: אבטחה, ניהול, אינטגרציה)
- תהליכי עבודה וניתוח שוטפים (למשל: דוחות, נגישות עצמית לניתוח אד-הוק, לוח מחוונים ניהולי)
- תהליכי ניתוח וחקר יצירתיים עבור Data Scientists ויכולת הטמעה שלהם בתהליכים שוטפים בארגון (למשל: ניתוח מידע לא מובנה, Visual Analytics, Machine Learning, Cognitive Analytics, R)
- יבמ תומכת בארגונים בכל רמה נדרשת על מנת לספק את הערכים הנדרשים לארגון, למנמ"ר ולמנהל ה BI:
- יחידת שירותי יעוץ עסקי - Global Business Services (GBS) לייעוץ ויישום פרויקטים מתקדמים
- כלי תוכנה אינטגרטיביים מהמובילים בעולם, על-מנת לספק פתרונות המותאמים לצרכים הנזכרים לעיל
- תשתיות ענן מתקדמות ביותר המאפשרות את ניהול פתרונות ה- BI והאנליטיקה המתקדמת על ענן מחשוב או SaaS.

יחידת שירותי הייעוץ העסקיים של יבמ

יחידה זו מיישמת פרויקטים רבים בארץ בצד האנליטי הארגוני, ביניהם פרויקטים מתקדמים ביותר הכוללים ניתוח מידע לא מובנה, ניתוח טקסט בעברית ובשפות נוספות, Machine Learning, Visual Analytics, Cognitive Computing ועוד. היא מציגה הוכחת ערך משמעותי וייחודי לפרויקטים בעלי המאפיינים החדשים (Visual, Text Analytics, Analytics וכו') עבור הארגונים בהם היא פועלת. לקוחות יכולים למנף את הניסיון והמומחים של יבמ מחו"ל, מעבדת המחקר בחיפה (השנייה בגודלה מחוץ לארה"ב), מעבדת התוכנה בישראל וצוות מומחים מקומי בעל הצלחות מוכחות, שילווה את הלקוח בדרך להצלחה עסקית מהותית (דוגמאות יינתנו לפי בקשה).

בתחום התוכנה

יבמ השקיעה מיליארדים רבים בשנים האחרונות ברכישה ופיתוח של כלי אנליטיקה מתקדמת.

להלן שימושים שונים של רכיבי תוכנה של יבמ הממחישים את היכולות השונות המוצגות בארכיטקטורה הכללית:

- InfoSphere Streams TM - תוכנה מעולם ה Real Time Analytics המאפשרת ניתוח של מידע מובנה ולא מובנה בזמן אמת בפלטפורמה סקלרית, לינארית לא מוגבלת. בנוסף התוכנה כוללת Toolkits לעולמות תוכן שונים וכן אינטגרציה מלאה עם סביבת יבמ (למשל, אינטגרציה מלאה עם SPSS Cognos, BigInsights, Moduler)
- InfoSphere BigInsights – תוכנה מבוססת Apache Hadoop המשווית לרכיב Data Exploration Archive בארכיטקטורה, הכוללת רכיבים ושיפורים רבים שאינם חלק מה Hadoop הסטנדרטי, כגון: מנוע ניתוח טקסט, מנוע Machine Learning, ממשק BigSQL לגישה סטנדרטית ב- SQL למידע. לתוכנה, קישוריות מלאה וסטנדרטית עם כלי יבמ האחרים בתחום האנליטיקה, כגון: SPSS, Cognos, Streams ועוד.
- Probabilistic Matching Engine-PME - הינו רכיב המאפשר אינטגרציה מעניינת וייחודית של תוכנה מעולם ה- MDM, המורץ מעל גבי סביבת Hadoop של יבמ ומאפשר יצירת רשומת לקוח אחודה על בסיס מודלים הסתברותיים שהוכחו, כטובים מאוד בעולם איחוד מידע מרשתות חברתיות שונות וכן איחוד מידע חיצוני ופנים ארגוני
- InfoSphere Watson Explorer – תוכנת אינדוקס וחפוש ארגוני המתאימה למידע מובנה ולא מובנה גם בסביבת Hadoop, גם בסביבת אפליקציות ארגוניות וגם למידע חיצוני בארגון
- SPSS Moduler – מודול מתקדם ליצירת מודלים שיכולים להשתלב גם בריצה בזמן אמת ב- Streams, גם בכלים כגון Unica וגם בצורה עצמאית על שרת אנליטי יעודי משפחת מוצרי SPSS השלמה נותנת מענה רחב ומקיף לכל נושא ה Data Mining (גם מעל Hadoop)

- קרי, ניתוח מידע מובנה ובלי מובנה (Text Analytics) באמצעות מגוון רחב ועשיר של מודלים אנליטיים ללמידה מבוקרת ובלתי מבוקרת. משפחת SPSS מנגישה את נושא ה-Data Mining לאנליסט העסקי באמצעות ממשק משתמש ידידותי ואינה מחייבת כתיבת קוד
- Information Catalogue - מאפשר למשתמשים לחפש/ לאתר נכסי מידע בקטלוג ארגוני (ריכוף מעל פריט המידע ע"מ לקבל קונטקסט ומשמעות, בחירה במידע הדרוש ויצירת Data Collection המכיל פריט מידע רצויים המתאימים לו, ומשם אינטגרציה של הנתונים, בקליק, במיקום חדש - בין שהוא במחסן הנתונים, תחת Hadoop או בענן)
- ישנם מספר כלים ופתרונות לעולם ה- Big Data and Smarter Analytics אשר יכולים לספק תובנות והזדמנויות עסקיות מסוגי מקורות המידע השונים ומהותם, תוך התחשבות במידת הזמן, המקור ונפחי המידע ההולכים וגלים, לדוגמא:
- ICA- IBM Content Analytics – סוויטה מלאה לניתוח תוכן וטקסטים הכוללת ממשק משתמש נח ונעים להפקת תובנות חדשות עבור הארגון
- i2 – כלי מחקרי מודיעיני שמאפשר יכולות ניתוח גרפיות מרשימות (Visual Analytics) למציאת קשרים (Link Analysis) ועוד.
- והרשימה עוד ארוכה אך לא נרחיב כאן...
- כדאי לציין, שיבמ מציעה כיום גם הצעת רישוי ייחודית: Forward looking BI, המאפשרת לבצע ניתוח נתונים, פרדיקציה ותכנון מעל דו"חות Cognos באמצעות פתרון המשלב כלי SPSS, Cognos, ו- TM1 ברישוי אטרקטיבי אחד. את כל היישומים האנליטיים, או את חלקם, יכולים ארגונים לפתח בענן. ליבמ פתרונות מהמתקדמים והמובילים ביותר בעולם בנושא זה:
- SoftLayer - מאפשרת הקמת אפליקציות ושירותים על ענן, אבל מאפשרת בנוסף יכולת הפרדת שרתים פיזיים, עד כדי הקמת שרתים בתוך הסביבה הארגונית של הלקוח כחלק מהענן (הצעה ייחודית ליבמ). בצורה זו, ניתן ליהנות מתשתיות ענן מבלי "להוציא" את המידע הארגוני מחוץ לתחומי הארגון
- IBM BlueMix - פלטפורמה של שירותי ענן המאפשרים הקמה של יישומים, במהירות הבזק, תוך שימוש ברכיבים מתוך קטלוג רחב היקף. בין היתר, ניתן למצוא שירותים (Services) מתחום האנליטיקה והניתוח, בקטלוג המוכן ליישום.

דוגמא מעניינת מהתחום, כאמור לעיל, הינה פתרון Counter Fraud של יבמ. הפתרון משלב מגוון כלי טכנולוגיים ב bundle עסקי אחד במחיר תחרותי המאפשר לכל ארגון להטמיע מערכת ניהול הונאות מהמתקדמות בעולם במהירות וביעילות. חלק מהכלים המשולבים בפתרון הנם:

- IBM Content Analytics (ICA)
- IBM Advanced Case Management (ACM)
- i2 Enterprise Analytics Platform (i2 IAP)
- FileNet™
- InfoSphere Streams™
- InfoSphere BigInsights
- משפחת מוצרי Cognos
- משפחת מוצרי SPSS

יישום הפתרון מבוצע באמצעות צוותים משולבים: מומחי תוכן מחו"ל בעלי ניסיון בעבודה עם המערכות בארגונים גדולים בעולם וצוות הייעוץ וההטמעה המקומי בחטיבת השירותים של יבמ ישראל.

אנשי קשר:

דוד בר – מנהל תחום Strategy & Analytics, GBS
דוא"ל: davidbar@il.ibm.com, טל.: 050-9165375

מיכל שני-פלר – מומחה מכירות בתחום Business Analytics
דוא"ל: michalsu@il.ibm.com, טל.: 050-505369178

איילת יוסף – מומחה מכירות בתחום Business Analytics
דוא"ל: ayelety@il.ibm.com, טל.: 050-2330623

תגובת חברת מיה מחשבים:

נשמע כי ארגונים רבים מתחילים להבין את הצורך בחיזוי אנליטי, ואת החשיבות של ללמוד מהמידע ההיסטורי הקיים בארגון על שצופן העתיד. אנו מבקשים לחלק את התייחסותנו ל- 3 חלקים מרכזיים:

א. הצורך הארגוני:

הארגונים מבינים את כוחו של המידע ומתחילים להבין את השלב הבא של ה-BI שהוא השלב האנליטי-חיזוי. נוצר הכרח בפעילות פרו-אקטיבית בכל פן עסקי, אם בשיווק, בכספים, כוח אדם, לוגיסטיקה ועוד. אנליטיקה מאפשרת למנף את תוצרי כלי ה-BI על ידי הבנה של השאלות העסקיות *הנכונות* שיש לשאלות. מנהלים מתחילים לדרוש מאנשי המידע לקבל כלים לחיזוי וניתוח מידע. בארגונים מתחילות להיעשות שתי פעולות הכנה מרכזיות: האחד, טיפול באיחוד ואיכות המידע – שלב בסיסי והכרחי בדרך לניתוח וחיזוי נכון (יכולות ETL, DQ וכו'). השני, מינוי מומחים שיהיו ה-Data Scientists של הארגון ויוכלו לשלב ידע סטטיסטי עם הבנת היכולות הטכנולוגיות וכמובן הבנת הפן העסקי של הארגון ומהו המידע העסקי של הארגון (ספציפי לכל ארגון).

ב. טכנולוגיה וכלים:

- כדי ליישם את "האמת הארגונית האחת" תוך שילוב עם כלי חיזוי, יש לעשות שימוש בכלים טכנולוגיים רבים, שרק השימוש בהם יאפשר להשיג יעדים אלו. להלן רשימה של כלים ויכולות מרכזיים והכרחיים:
- יישום הפתרון באמצעות שילוב של כלי חיזוי שונים שיתנו יחד את הפתרון השלם (כלי אופטימיזציה, ניתוח Mining, forecasting, תמחור, סיכונים, ניהול קמפיינים וכו').
- שילוב של כלי Self Service עם כלים למקצועי חיזוי וסטטיסטיקאים
- Bigdata – עבודה עם בסיסי נתונים המאפשרים לטפל בכמות אדירה של נתונים, מידע מסוגים שונים, מידע איכותי, ובמהירות. יכולת עבודה עם כלים דוגמת Hadoop בסיסי נתונים NoSQL ובקיצור – כל בסיס נתונים.
- עבודה In-Memory (על גריד של שרתים ולא בשרת בודד) ויכולת לנתח במהירות רבה כמויות אדירות של מידע.
- כלי BI מתקדמים המשולבים בכלי Office ומחשבים ניידים וטלפונים.

ג. דוגמאות לישומים ופתרונות אנליטיקה הנפוצים כיום הם:

1. ניתוח ערוצי דיגיטל – חווית לקוח והעצמת ערך הלקוח לארגון
2. ניתוח רשתות חברתיות
3. Sentiment analysis
4. כריית מידע של נתונים טקסטואליים
5. אופטימיזציה תהליכים – פעילויות שיווקיות, תהליכי גבייה, מלאים ועוד
6. חיזוי ביקושים
7. סיכונים
8. תמחור
9. אופטימיזציה מלאי ורכש
10. סייבר אנליטי

חברת SAS, עפ"י גרטנר, עושה למעלה מ-36% מהאנליטיקה בעולם. ככזו, לחברה מגוון רחב ביותר של מוצרים בתחומי: Data Management, אנליטיקה, BI. מוצרים SAS פועלים על כל מערכות ההפעלה מ-MF ועד PC, ועל כל בסיסי הנתונים הקיימים בשוק. כמו כן, מוצרי החברה פועלים בטכנולוגיה ייחודית של SAS בשם High Performance Analytics ותומכים בסביבת BigData ו-In Memory המאפשרים טיפול בכמויות מידע ענקיות ובמהירות רבה.

על בסיס המוצרים הנ"ל ותשתיות ה-High Performance Analytics נבנו פתרונות הכוללים Best Practices למגזרים השונים (ביטוח, בנקאות, ממשלה, בריאות וכו'). SAS מיוצגת בישראל ע"י מיה מחשבים ולה כ-70 עובדים, מעל 50 נמצאים אצל לקוחות כיועצים – מומחים.

איש קשר: מוטי סדובסקי, מיה מחשבים. טל: 054-3385118, אימייל: motti.sadovsky@sas.com

תגובת חברת HP:

יותר ויותר ארגונים היום מבינים שהכלים המסורתיים בסביבת מחסן הנתונים נכונים לצרכים הסטטיים של הארגון, ומאוד מוגבלים ביכולת שלהם לתת מענה ל:

- ניתוח עמוק של המידע
- התמודדות עם צורכי ניתוח ואופטימיזציה בזמן אמת
- יכולת לייצר עניין אצל לקוחות דרך הבנת הפרופיל מתוך ניתוח כמויות מידע היסטורי גדולות
- ועוד

חברות טכנולוגיה רבות בארץ כבר מבינות את היתרון שבשימוש בכלי קיבול אנליטי כמו ורטיקה ומטמיעות אותו כחלק מהסביבה שלהן, בין אם לצורכי ניתוח אנליטיים לאנליסטים בארגון ובין אם כחלק ממוצר כולל אשר נבנה עבור לקוחות של החברה (OEM).

מי שמימש פתרון זה מימש קפיצה אדירה וקיבל למשל שיפר ביצועים שבין פי 50 לפי 1000 ואף יותר מכך.

עצם העובדה שורטיקה הינו פתרון מבוסס SQL סטנדרטי, עם זמן קצר ביותר למימוש ההטמעה, הופכת את ההחזר על ההשקעה למהיר מאוד ועלות בעלות כוללת נמוכה מאוד.

מרגע שהארגון נחשף ליכולות, מתנסה בביצועים תוך הוכחת יכולת קצרה, מתקבלת ההבנה שהשינוי הינו הכרחי לצורך היכולת לשמור על היתרון הטכנולוגי וההובלה העסקית.

איש קשר:

ליאור צברי, מנהל תחום VERTICA ב- HP. טל: 052-4840891

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

אינדקס לפי שם מחבר (כרך א' עד כרך כא' 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 יכין	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 וחילוץ ישויות בסביבות שונות מבוסס ניתוח מורפולוגי	ארנן ליאור גילהר חיים	כא'	2	12/2014
חיפוש ארגוני, סיכום מפגש שולחן עגול	בדווגין ליה	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 בידעיות 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	גיורא שמעוני	יא'	1	1/2005
מורפולוגיה למנועי חיפוש	גיורא שמעוני	יד'	1	1/2008
ישום GSA במאגר פסקי דין	גיורא 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 ליאונד וויסיפון אורן	יז'	1	1/2011

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
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
הבנת שפה ביישומי טקסטים	דרי' דגן עידו	יג'	2	6/2007
טכנולוגיית מיקרוסופט לאחזור מידע (ביג דטה)	דוד מאור	כ'	1	6/2013
הוצאה לאור אלקטרונית	דננברג אמיר	ג'	1	2/1997
הסבת ארכיון צה"ל ממחשב WANG למחשב HP תחת UNIX	דננברג 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
		כא'	1	6/2014
		כא'	2	12/2014
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

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
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 ענין	דרורי עפר	ב'	1	5/1996
עיצוב ממשק משתמש במערכות מידע	דרורי עפר	ה'	1	1/1999
שילוב מערכות אחזור טקסט ומערכות מידע קונבנציונליות	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	ד'ר' הנדל רות	ב'	2	11/1996
חברי הכנסת כצרכני מידע	ד'ר' מרקוס רבקה	יט'	2	12/2012

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
עיבוד שפה טבעית בערבית	הרשברג נמרוד כפיר בר	יד'	1	1/2008
כלי לחילוץ משמעויות מטקסט	ובר שמואל של-בר עוז	כא'	2	12/2014
עיצוב ממשק משתמש ל- WEB	ווידנפלד צביקה	ד'	2	5/1998
כיווני פיתוח באחזור טקסט TQL	וייזל גלעד	אי'	1	4/1995
רפורטואר, מערכת שאילתות מותאמת ללקוח	וייזל גלעד	כא'	2	12/2014
חיפוש מידע מבוסס ניתוח טקסט	וינהבר אורי	יב'	1	1/2006
ניתוח שאילתות באתרי מכירות	דר' ולן אינגריד	יג'	2	6/2007
מחשוב המתווה הלכסיקוגרפי 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
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 expetnext	ד"ר חנני אורי	א'	1	4/1995
מנתונים לידע - MindCite	ד"ר חנני אורי	טי'	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	טי'	1	1/2003
חיפוש טקסט מלא בעברית ובערבית - הבעיה והפתרון	ירדני לאורה	י'	1	1/2004
תכונות מוצר לניהול הידע D2K.NET	כהן אייל	טי'	1	1/2003
חיפוש תלוי הקשר, עקרונות ודוגמאות	כהן חנן	חי'	1	1/2002
זכרון לטווח רחוק	כהן שגיא	יט'	2	12/2012
אחזור מסמכים משובשים	לבנה משה	חי'	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 סינית! כלים לניתוח טקסט בעברית ובשפות אחרות"	ליטוב דוד אנדרו גולדמן	יז'	1	1/2011
המרת אתרים מעברית ויזואלית ללוגית	ליס אורלי	טי'	2	6/2003
ארגון תוצאות חיפוש ממנועי אחזור באינטרנט	דר' לסט מרק	חי'	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 ממוחשב	מודן דורון	כ'	2	10/2013

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
תכונות מנוע החיפוש WizDoc	מידן אברהם	ט'	1	1/2003
WizDoc - מנוע חיפוש לפי משמעויות בעברית ובאנגלית	מידן אברהם	ט'	2	6/2003
חיפוש לפי משמעויות	דר' מידן אברהם	יד'	1	1/2008
תכונות מנוע החיפוש Fast	מימוני אלון	ט'	2	6/2003
ארכיטקטורה של מנוע החיפוש Fast Search	מימוני אלון	ט'	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	מרגלית ששון	יג'	2	6/2007
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
חיפוש OPEN SOURCE במאגרים גדולים (BIGDATA) עם מנוע החיפוש לוסין ו-Elasticsearch	סין-הרשקו איתמר	כא'	1	6/2014
עידון תהליכי חיפוש במאגרי מידע והצגתם למשתמש	סליטר 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
גממות 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
BI במימוש עצמי, שרותי תחקור לאנליסט בסביבת ביג דאטה	פסטרנק רועי	כא'	2	6/2014
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

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
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 תחבירי חלקי	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
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
Context - מוצר לניהול תוכן	שושני יניב	יב'	1	1/2006
מערכות BI לקידום יעדי הארגון	שחור אסף	כא'	2	6/2014
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
יישום GSA במאגר פסקי דין	שמעוני גיורא	יט'	2	12/2012
מנועי חיפוש ארגוניים	שמעוני עינת	טו'	1	1/2009
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
מגמות בתחום ניהול ידע	שמעוני עינת	כא'	1	6/2014
סיכום מפגש שולחן עגול, ניתוח טקסט	שמעוני עינת	כא'	2	12/2014
סיכום מפגש שולחן עגול - אנליטיקה	שמעוני עינת	כא'	2	12/2014
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	כ'	2	10/2013
XML - המסלול המהיר לכלכלה החדשה	שרוטר גרט	ז'	2	6/2001

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
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
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

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

אינדקס לפי שם מאמר (כרך א' עד כרך כא' 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 גיימיל	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	ד"ר 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	ד"ר 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	רן גוראון	ד'	1	2/1998
היפרטקסט וקבלת החלטות				

המשך אינדקס לפי שם מאמר (כרך א' עד כרך כא' 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 טקסנומיה ארגונית בשילוב ממשק LCC&K	ויקי מרדכי	יג'	2	6/2007
הצגת תוצאות מקובצת של תוצאות חיפוש עפ"י טקסנומיה ארגונית בהתבסס על ממשק LCCK	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	אלינור בנחקון	ה'	2	5/1999
ויזואליזציה של מידע בהתבסס על קשרים מובנים וניתוחם	פבל חרדק	טז'	2	6/2010
ויזואליזציה של תוצאות חיפוש במערכות אחזור מידע	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	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 כא' 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
		כא'	1	6/2014
		כא'	2	12/2014
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 וחילוץ ישויות בסביבות שונות מבוסס ניתוח מורפולוגי	ליאור ארן חיים גילהר	כא'	2	12/2014
חיפוש חופשי בטקסטים סרוקים בעברית	גורא 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
חיפוש OPEN SOURCE במאגרים גדולים (BIGDATA) עם מנוע החיפוש לוסין ו-Elasticsearch	איתמר סין- הרשקו	כא'	1	6/2014
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	גורא שמעוני	יט'	2	12/2012

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
כיווני פיתוח באחזור טקסט TQL	גלעד ויזל	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 לחילוץ משמעויות מטקסט	שמואל ובר גלעד ויזל	כא'	2	12/2014
כלי חיפוש סטנדרטיים במערכות ומוצרים קיימים	מוטי חגיז	ב'	2	6/2006
כלי לחיפוש שיתופי באינטרנט - Antworld	שפירא ברכה	ז'	2	6/2001
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 בתחום ניהול ידע	עינת שמעוני	כא'	1	6/2014
מגמות עתידיות בעולם איחזור המידע expetext	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	טו'	1	1/2003
מנוע החיפוש RetrievalWare	נחמה אנדלמן וצבי קמר	טו'	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	חי'	2	6/2002
מנוע חיפוש כתשתית לאוטומציה של תהליכים ידניים במאגרים טקסטואליים	מיקי קולקו	ה'	1	1/1999
מנוע חיפוש לשפה העברית	עפר דרורי	טו'	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	חי'	2	6/2002

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
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 (גרסה 7.2009)	דרורי עפר	טז'	1	1/2010
מנועי אחזור טקסט בעברית - רשימת ספקים (גירסה 10.2010)	דרורי עפר	יז'	1	1/2011
מנועי חיפוש ארגוניים	ענת 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 רונן פלדמן	יב'	1	1/2006
מערכת נוהלים בטכנולוגיית אינטרנט-נט	איציק הוך	ד'	2	5/1998
מערכת BI לקידום יעדי הארגון	אסף שחור	כא'	2	6/2014
מפת הדרכים של החיפוש הארגוני במיקרוסופט - MOSS	רונה לוסיג	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 הצל"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 שאלות באתרי מכירות	ד"ר אינגריד 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 מפגש שולחן עגול, ניתוח טקסט	ענת שמעוני	כא'	2	12/2014
סיכום מפגש שולחן עגול –אנליטיקה	ענת שמעוני	כא'	2	12/2014
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

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
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 שאלות 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
רפורטאר, מערכת שאלות מותאמת ללקוח	גלעד ויזל	כא'	2	12/2014
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
רשמים משולחן עגול בנושא Web 2.0 לצורך ניהול ידע ארגוני	ענת 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	ענת 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	ליאניד 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 סלסנס	טו'	2	6/2009

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

שם המאמר	שם המחבר	כרך	גליון מס.	תאריך
A Modular Information Extraction System	יזהר רגב מאיה גורודצקי רונן פלדמן	יב'	1	1/2006
BI במימוש עצמי, שרותי תחקור לאנליסט בסביבת ביג דאטה	רועי פסטרנק	כא'	2	6/2014
Context - מוצר לניהול תוכן	יניב שושני	יב'	1	1/2006
Full Text - מנוע חיפוש בעברית	שחר אורון	ז'	1	1/2001
Guidance : כלי ליצירת ממשק גרפי למערכת MF	דורר 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
MKknowledge - תוכנה לקבלה, ניתוח והצגה של ידע	ניר לוי	יז	1	1/2011
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