



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 לדין 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 מישהו 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 כמותני? 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 "מטעם" 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 הBI/אנליטיקה - מעבר משאלות מובנות (Queries) לגילויים (Discoveries) ושיטוטים במידע. כשעושים סקר מחקר "סטנדרטי" ישנן שאלות שלא נחשוב לשאול אותן. לעומת מחקר הקשבה - מישהו 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, מייעצות, 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, ייהנה מאינפוסטים 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 הבעייתית, 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 מציינים 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 התחרותי 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 אמת 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 לידיים לצורך ספציפי (לדוגמה, מידע על "אירועים" בחיי לקוחות שיכולים להוות ליד – לדוגמה, מעוניין לרכוש דירה ולכן כדאי להציע לו ביטוח משכנתא בדיוור הקרוב). זיהוי לידיים 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 classification או auto categorization) 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

edval@matrix.co.il , 052-2385946

תגובת חברת IFN-יעל:

1. אכן קיימת דרישה הולכת וגוברת של לקוחותינו לבצע חיפוש טקסט בכלל מאגרי הארגון ולא רק במסמכים. לדוגמא, דוא"ל וצרופותיו, 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:

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