**פרוייקט גמר**

StarTrace

**מחברים**: נטע לי להט, גל בן יאיר, יונתן אברהמי ועמית וייל

**מוגש לבית הספר למדעי המחשב**

**המסלול האדמאי, המכללה למנהל**

**כחלק מהדרישות האקדמאיות**

**לתואר B.Sc. במדעי המחשב**

אב התשע"ה ראשל"צ יולי 2015

**פרוייקט גמר**

StarTrace

**מחברים**: נטע לי להט, גל בן יאיר, יונתן אברהמי ועמית וייל

**המנחה המאשר**: ד"ר משה בוטמן

**בית הספר למדעי המחשב**

**המסלול האדמאי, המכללה למנהל**

אב התשע"ה ראשל"צ יולי 2015

**תוכן העניינים – לערוך בסוף**

[תקציר - 5 -](#_Toc392447285)

[הקדמה - 6 -](#_Toc392447286)

[הצדקה אקדמאית - 6 -](#_Toc392447287)

[מטרת הפרויקט - 6 -](#_Toc392447288)

[גוף הפרוייקט - 7 -](#_Toc392447289)

[1. ארכיטקטורה - 7 -](#_Toc392447290)

[2. מבנה בסיס הנתונים - 9 -](#_Toc392447291)

[3. עבודה מול ה Team Foundation Server של מיקרוסופט - 10 -](#_Toc392447292)

[אפיון המסכים - 13 -](#_Toc392447293)

[1. מסכי המערכת - 13 -](#_Toc392447294)

[מסך ניהול צוותים וחברי צוות - 13 -](#_Toc392447296)

[מסך ניהול תכונות ותתי תכונות - 14 -](#_Toc392447298)

[מסך ספרינטים - 15 -](#_Toc392447301)

[מסך צפייה בלו"ז צוותי - 16 -](#_Toc392447307)

[2. שימוש במסכי המערכת - 17 -](#_Toc392447310)

[טכנולוגיות בשימוש - 18 -](#_Toc392447320)

[קשיים בדרך ופתרונם - 20 -](#_Toc392447321)

[אפשרויות יישום לצורך הרחבת המערכת - 20 -](#_Toc392447322)

[האלגוריתם - 21 -](#_Toc392447323)

[מסקנות, יתרונות וחסרונות - 33 -](#_Toc392447324)

[תרשימים - 34 -](#_Toc392447325)

[מקורות - 35 -](#_Toc392447330)

[שימוש בספריות Open Source - 35 -](#_Toc392447331)

תקציר

בעידן הרשתות החברתיות וכמויות המידע העצומות הקיימות ברשת ניתן לצבור מידע רב היכול לשמש להצלחת שיווק ופרסום בעולם כולו.

לפני תחילת מסע פרסום ניתן להשתמש במידע זה ככלי על מנת לדעת מה דעת הרוב על מפורסם בעיר מסוימת ובכך למנוע כישלון ובזבוז כספים על פרסום לא מוצלח.

StarTrace תעזור לבחור במפורסם המתאים ביותר למסע פרסום במיקום מסוים (עיר או מדינה) בעולם בהתבסס על ניתוח סנטימנטלי של תגובות משתמשים על אותו המפורסם ברשתות חברתיות.

בממשק למשתמש יש להזין שם של מפורסם ועיר ולאחר חישוב יוצגו למשתמש אחוזי הדעות החיוביות, השליליות והניטרליות על אותו המפורסם בעיר שנבחרה.

ממשק ה- WEB ימומש ע"י טכונלוגית AngularJS ביחד עם חבילת עיצוב Bootstrap, ע"מ ליצור מראה פשוט ונקי. שרת האפליקציה ישלח בקשות אל שרת מכונת הלמידה, אשר תנתח אותן לפי אלגוריתמי הניתוח הסנטימנטלי, ויקבל חזרה את תוצאות הניתוח. מכונת הלמידה תאסוף את המידע באמצעות הממשקים השונים שמציעות הרשתות החברתיות השונות (טוויטר, פייסבוק, אינסטגרם וכו'). הקוד של מכונת הלמידה ימומש בPython 2.7

הקדמה

כיום ברשת האינטרנט קיים מידע רב אודות מפורסמים ומעריציהם. המידע אודות המפורסמים מגיע בכמויות גדולות וממקורות שונים ומגוונים. לאיסוף ועיבוד המידע פוטנציאל עסקי רב.

הגדרת עולם הבעיה - לפחות עמוד

הגדרת הבעיה ומה נעשה בתחום - 2 עמודים לפחות

פתרון - הסבר על הפתרון המוצע וכיצד שונה מאחרים

הצדקה אקדמאית

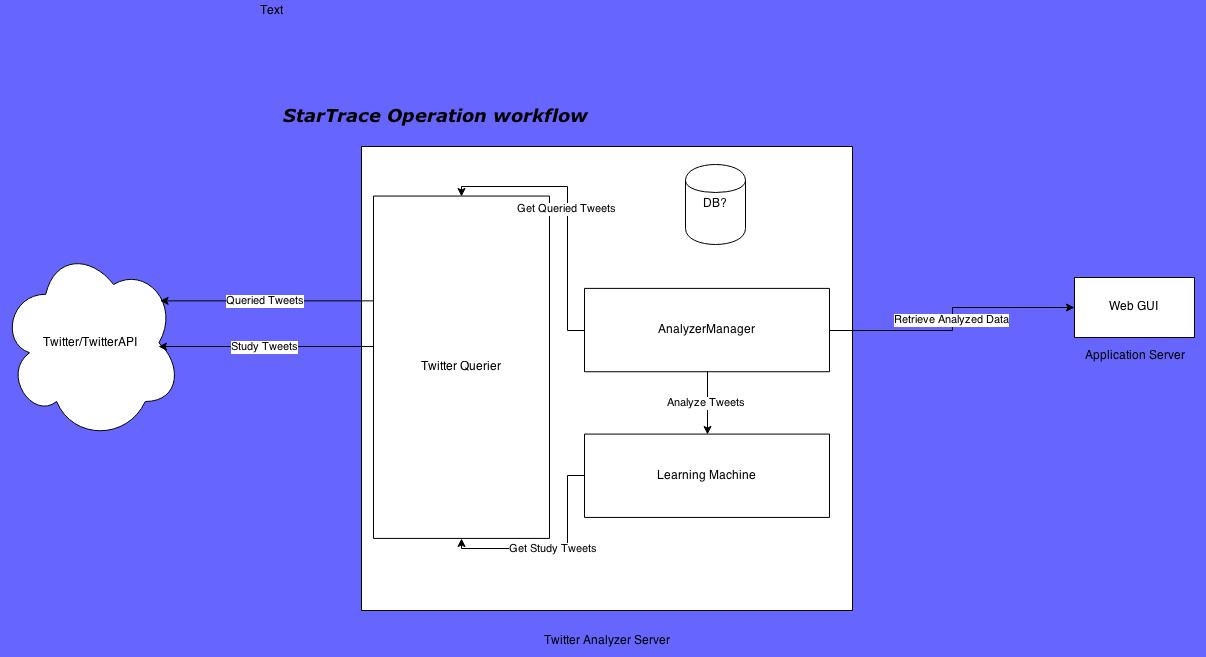
כדי להסיק תובנות בעלות פוטנציאל עסקי מתוך מאגרי הנתונים הנאספים מרשתות חברתיות על מפורסמים נשתמש באלגוריתמי למידה אשר נלמדו בקורס למידה חישובית וכריית נתונים.

מטרת הפרויקט

מטרת הפרויקט היא מציאת תובנות עסקיות בעזרת אלגוריתמי למידה על בסיס מידע רב שנאסף מרחבי האינטרנט. כדי לקבל החלטות עסקיות בתחום הפרסום נדרש להשתמש במידע של דעת הקהל הרחב על מפורסמים כדי שתהליך הפרסום יהיה מוצלח. נבנה מערכת שמספקת ידע לגבי דעת הקהל על מפורסמים על פי ערים בעולם על בסיס מידע..

תרשים בלוקים\מודולים כללי של המערכת

יוני – זכור לי שעשית פעם תרשים של הארכיטקטורה, נראה לי שצריך לשפץ אותו ולהוסיף לו עוד תרשימים



DR

SRS

מתודולוגיות וכלי פיתוח, חומרה ותוכנה

המערכת מחולקת לשני חלקים – צד שרת וצד לקוח. צד השרת פותח לחלוטין ב-Python על מנת להשיג זריזות וגמישות בעבודה עם הממשקים. על מנת להפוך מיקום משם של מקום לקואורדינציות השתמשנו בשירות רשת של הכנס שם כאן. את ממשק ה-REST פיתחנו באמצעות טכנולוגייה מבוססת Python בשם "הכנס שם כאן". נטע לי תרחיבי על צד לקוח

המערכת - כולל תמונות מסך של קלט ופלט המערכת

משימות על פי זמנים והסבר על עמידה בזמנים

* בדה"ת לטכנולוגיות ודרכי איסוף נתונים מרשתות חברתיות שונות- ינואר 15
* בחירות טכנולוגיות ותחילת מימוש – פברואר 15
* פיתוח אופן הלמידה ובחינת פלט האלגוריתם – מרץ 15
* פיתוח ממשק למשתמש – אפריל 15
* הכנת תקצירים, ברושור ופוסטר לפרויקט - מאי 15
* הצגת הפרויקט – יוני 15
* ספר פרויקט – יולי 15

צורת העבודה שבה עבד הצוות למימוש הפרויקט הייתה חלוקת משימות בין חברי הצוות כדי שנעמוד בלוח הזמנים הנדרש לפרויקט, עמדנו בזמני העבודה שנדרשו.

קשיים במהלך הפרויקט וכיצד נפתרו

הקושי המרכזי במהלך הפרויקט היה למצוא API ידידותי אשר יאפשר לנו להוציא מידע על אנשים מרחבי הרשתות החברתיות. ראשית ניסינו לעבוד מול Facebook, אבל לאחר שגילינו שהם ביצעו שינויים ב-API שלהם שמונעים הוצאת מידע בסיסי על משתמשים החלטנו לעשות שינוי כיוון (Pivotting) ולהוציא מידע מ-Twitter במקום. כעת נתקלנו ב-API ידידותי יותר המאפשר לאחזר Tweetים ממשתמשים שונים ובנוסף לקבל Tweetים לפי מיקום. לאחר שאחזרנו את ה-Tweetים כל מה שנותר הוא לנתח אותם ולייחצן את התוצאת באמצעות ממשק REST מעל HTTP. קושי נוסף שנתקלנו בו הוא שה-API של Twiiter מאפשר להוציא רק מספר מסוים של Tweetים מתקופה שלא תעלה על חודש. פתרון שחשבנו על מנת לפתור את הבעיה הזו היא לשמור את הTweetים במסד נתונים משלנו, ובהמשך ליצור ממשק המבצע שילוב בין אחזור מידע מ-Twitter לבין עבודה מול מסד הנתונים שלנו על מנת להפיק ביצועי למידה טובים יותר ולהגיע לתוצאות איכותיות יותר.

ביבליוגרפיה