**צ'אט-בוט עזר לתכנות. Chat-Bot Programing Helper.**

שם המגיש: איתי עזרא Name : Itai Ezra  
ת.ז: 026616128 ID:

שם המנחה: רועי יצחק Royi Itzhak Instructor Name:  
חתימת המנחה: \_\_\_\_\_\_\_\_\_\_

תאריך: 15.02.2014

**תקציר:**

עד היום בכלל לא למדתי או התעסקתי בבינה מלאכותית.

אני נכנס לפרויקט עיור אך עם המון אמונה ורצון טוב. אני רואה בבינה מלאכותית וברובוטיקה את העתיד.

בקצרה הרעיון שלי הוא לעזור למתכנתים מתחילים למצוא פתרונות לבעיות פשוטות כמו, בניית מערכים, מימוש לולאות, הדפסה למסך וכו'.

כמובן שעם יותר זמן ומשאבים יהיה אפשר לפתח את האפליקציה למצוא פתרונות לבעיות מסובכות יותר.

המטרה היא לחסוך מהמתכנת הצעיר לחפש את הפתרון שלו באינטרנט ולבחוש בכל הפורומים והאתרים השונים עד שהוא ימצא את הפתרון, ובמקום להציע לו אפליקציה זמינה שתמצא לו פתרון יעיל בזמן סביר.

**תיאור מסגרת הפרויקט:**

האפליקציה תהיה צ'אט בין המשתמש למחשב.

המשתמש יוכל לשאול את המחשב ולבקש את העזרה שלו בפתרון בעיות בתכנות (C#).

האפליקציה, בעזרת בינה מלאכותית, תנסה להבין את הקלט מהמשתמש ולמצוא פתרון מתאים לבעיה.

למעשה יש שני חלקים לאפליקציה:

1. סריקה של פורומים ידועים וגדולים על מנת להוציא מהם מאגר גדול ככל האפשר של שאלות ותשובות ולבנות מהם מאגר מידע מסודר עם קשרים בין מילות מפתח על מנת לקשר בין בקשת המשתמש לתשובה הרלבנטית ביותר. בעזרת חישובים סטטיסטים האפליקציה תמצא לפי מספר מופעים של מילות מפתח בשאלה מה היא התשובה עם ההסתברות להיות המתאימה ביותר.
2. אפליקציית צ'אט בין המשתמש למחשב שבעזרת Keyword Extractor יקשר בין מילות מפתח מבקשת המשתמש לתשובה המתאימה ביותר ממאגר המיידע.

השאיפה היא להגיע למצב שהאפליקציה היא ממש plug-in ל Visual Studio והצ'אט יהיה בדיבור ולא בכתב, אך מפאת חוסר זמן והרצון לעמוד בזמנים, האפליקציה תעמוד בפני עצמה ותתבצע בכתיבה.

כמו כן אנחנו נצא מנקודת הנחה שהמשתמש שואל שאלות רלבנטיות לתכנות ובאנגלית בלבד.

עיקר העניין של הפרויקט הוא הבינה המלאכותית, היכולת להבין את הבקשה של המשתמש ולמצוא לו את הפתרון המתאים ביותר מתוך מאגר המיידע.

**תיאור הבעיה:**

הבעיה אותה אנחנו רוצים לפתור היא הצורך הבלתי נפסק של כל מתכנת, ובעיקר מתכנתים מתחילים, לחרוש את האינטרנט בכל פעם שהוא נתקל בפעולה שאותה הוא רוצה לבצע והוא לא יודע איך מפאת חוסר היכרות עם השפה או עם אלגוריתמים מתאימים.

מבחינת הנדסת תוכנה, ישנם מס' בעיות.

1. רוב הפורומים לא מספקים API ו/או את היכולת לגשת בצורה נוחה למאגר השאלות והתשובות שלהם.
2. רוב השאלות והתשובות בפורומים הם לקטע קוד ספציפי ולא שאלה כללית, מה שמקשה מאוד למצות מהם את השאלה הכללית והתשובה שלה.
3. בניית AI שהוא למעשה יהיה הקשרים בין מילות מפתח בשאלה לתשובה הכי מתאימה.
4. יצירת אינטראקציה אנושית ככל הניתן בין המשתמש למחשב.
5. בניית האפליקציה באופן כזה שבו היא תוכל להתגבר על בעיות כמו, אי מציאת פתרון, מציאת פתרון לא נכון.

**תיאור הפתרון (לבעיות התכנות):**

כאמור עדיין אין לי פתרונות לכל הבעיות.

1. שימוש במספר מצומצם של פורומים. עדיין המאגר אמור להיות גדול מספיק. אם אפשרות של הוספה "ידנית" למאגר.
2. שימוש Keyword Extractor כדי למצות את השאלה הכללית והתשובה שלה.
3. שימוש באלגוריתם בייסיאני לבניית ויצירת הקשרים ביין מילים מסוימות לתשובה מהמאגר. כלומר, ניקח את מילות המפתח מהשאלה, ולפי מספר המופעים של מילות המפתח בשאלה נמצא את התשובה עם ה"סיכוי" הגבוה ביותר להיות התשובה המתאימה.
4. יצירת מאגר של שאלות ותשובות ראשוניות על מנת לדלות מידע בסיסי מהמשתמש אך גם כדי ליצור אינטראקציה "אנושית" בסיסית.
5. ישנם מספר אפשרויות לפתור בעיה זו. נכון לעכשיו לאחר כל מתן פתרון המחשב יבקש פידבק מהמשתמש כדי לדעת אם הוא קיבל את הפתרון המתאים.

**פתרונות בספרות:**

כיום ישנם מספר רב של צ'אט בוטים עם יעודים שונים.

לרוב הטלפונים החכמים קיימות מס' אפליקציות של צ'אט בוטים, רובם משמשים על תקן "מזכירה": תזכורות, עזרה בהתמצאות, מציאת חנויות, מסעדות וכו'.

מספר רב של אתרי אינטרנט משתמש בצ'אט בוטים כדי ליצור אינטראקציה ראשונית עם המשתמשים באתר, לקיחת פרטים והמתנה לנציג שירות שימשיך את השיחה.

מחיפושים שבצעתי לא מצאתי תוכנות שיש להם עזרה בצורה של צ'אט-בוט, אני מאמין שיש כאלה אך השימוש בהם עדיין אינו נרחב ואני מאמין שהדבר הזה ישתנה בעתיד יחד עם התפתחות הבינה המלאכותית.

ניתן ליצור צ'אט-בוטים מהרמה הפשוטה ו "הטיפשה" ביותר עם מס' של שאלות (IF ELSE) ,

אפשר להשתמש במאגר מידע גדול ממנו הבוט ישלוף את התשובות שלו,

ואפשר להשתמש באלגוריתמים לומדים.

כאמור אין דרך אחת או שפה אחת כדי ליצור צ'אט בוט, הסיבה שבגללה בחרתי את הרעיון הזה היא כי הוא עדיין מתפתח ואני מאמין שברובוטיקה ובינה מלאכותית השמיים הם הגבול.

כאמור ישנם בשוק מספר רב מאוד של צ'אט בוטים, מפשוטים וטיפשים מאוד לאינטליגנטים במיוחד.

הנה מספר דוגמאות :

1. [אליס.](http://www.alicebot.org/)
2. [סירי](http://www.apple.com/ios/siri/).
3. [אלייזה.](http://he.wikipedia.org/wiki/%D7%90%D7%9C%D7%99%D7%99%D7%96%D7%94)
4. [מיטצוקו](http://www.mitsuku.com/).

מהדוגמאות האלה סירי היא האפליקציה היחידה שמציעה עזרה בתחום מסוים.

אלייזה הוקמה כפרודיה לטיפול פסיכיאטרי ראשוני בו המטפל לא ממש מכיר את המטופל ולכן לרוב עונה לשאלות המטופל בשאלה. אליס היא אפליקציית צ'אט בוט מעט הומוריסטית שהוקמה בהשראת אלייזה.

לא מצאתי אף צ'אט בוט לעזרה בתכנות בפרט או ככלי עזר לתוכנה קיימת כלשהי בכלל.

**תיאור אב טיפוס שמומש:**

כרגע המערכת עדיין לא באינטגרציה בין כל החלקים שלה.

ישנו צ'אט שעובד אך הוא עדיין חסר שימוש. כתבתי תוכנה נפרדת ששולפת מידע מפורומים, נכון לעכשיו אני עובד עם <http://stackoverflow.com/> מאחר ויש להם API נוח מספיק.

התוכנה שולפת מהמאגר שלהם את כל השאלות והתשובות.

הבעיה היא שהקשר היחיד בין השאלות לתשובות הוא מספר סידורי שלהם, ולכן כתבתי תוכנית נוספת שעושה סדר. התוכנית לוקחת כל שאלה ומוצאת לה את התשובה המתאימה עם הדירוג הגבוה ביותר ושומרת אותה בקובץ TXT נפרד. קבצי TXT אלה יהיו מאגר המידע של האפליקציה.

בעיה נוספת שנתקלתי בה היא שכמעט כל השאלות והתשובות בפורום הם לקטעי קוד ספציפיים, מה שמקשה מאוד לפענח מהם את השאלה הכללית והתשובה שאותה נרצה לתת למשתמש (עדיין לא התגברתי על בעיה זו).

חלק נוסף שכבר מומש הוא KeyWordExtractor שנכון לעכשיו יודע למצוא בקבצי TXT את המילים החשובות בקטע.

מבחינת תכנון, ה KeyWordExtractor יצטרך לעבור על כל קבצי הTXT ויחד עם תוכנה נפרדת לבנות את כל הקשרים בין המילים החשובות לתשובות שלהם. בנוסף הוא יהיה זה שישלוף את המילים החשובות מהצ'אט עם המשתמש כדי למצוא לו את התשובה.

למעשה בסופו של דבר השימוש במאגר יהיה כדי לבנות את הקשרים ביין מס' המופעים של מילות מפתח בשאלה, לתשובה המתאימה ביותר. ולאו דווקא להשתמש בתשובות עצמם מהמאגר.

**בדיקות של המערכת:**

* עדיין לא מתבצעות בדיקות מיוחדות של המערכת.

**מסקנות מהפתרון והמימוש:**

כאמור הפתרון עדיין לא הושלם.

* כאשר מתעסקים בבעיות בסיסיות/פשוטות בתכנות, אולי יהיה צורך בלהכניס מספר שאלות ותשובות למאגר באופן ידני, מאחר ורוב הפורומים לא מתעסקים בשאלות אלה.
* האפליקציה תענה על שאלות ב C# בלבד, אך נכון לעכשיו המאגר שצברתי הוא של כל השאלות והתשובות של <http://stackoverflow.com/> ולא רק אלה בשפת C#. יכול להיות שיהיה צורך למצוא פורום חלופי שמתעסק ב C# בלבד כמו MSDN, אך עדיין לא מצאתי כזה שיש לו API שניתן לעבוד איתו.
* יש צורך לכתוב אלגוריתם על פי המודל הבייסיאני.

**לו"ז וסיכוני המשך:**

**לו"ז:**

|  |  |  |
| --- | --- | --- |
| # | שלב בפרויקט | זמן משוער (ימים) |
| 2 | האפליקציה יכולה לעזור לי להדפיס "HELLO WORLD" | 3 |
| 3 | הטמעה של האלגוריתמים בתוכנת הצ'אט.(AI) | 7 |
| 4 | האפליקציה יכולה לעזור לי ליצור לולאה. | 3 |
| 5 | האפליקציה יכולה לעזור לי ליצור מערכים. | 3 |
| 8 | יצירת משחק איקס עיגול תוך שימוש בלולאות, מערכים ואובייקטים תוך שימוש מקסימלי בעזרה מהאפליקציה. | 7 |
| 10 | שיפור יכולות ואנושיות האפליקציה. | בזמן הנותר.. |
| 11 | סה"כ | 25+ |

**סיכונים:**

* אי מציאת תשובה לשאלה למרות שהיא קיימת במאגר.
* החזרת תשובה שלא תואמת את השאלה.
* החזרת תשובה אחרי פרק זמן ארוך מידי.
* חוסר אנושיות של האפליקציה.
* אי עמידה בזמנים.

**רשימת ספרות:**

* [KeyWordExtractor](http://ymatsuo.com/papers/ijait04.pdf)
* [StackOverFllow API](Stack%20OverFllow%20API)
* [Automatic Detection of Causal Relations for Question Answering](http://delivery.acm.org/10.1145/1120000/1119322/p76-girju.pdf?ip=79.177.131.248&id=1119322&acc=OPEN&key=BF13D071DEA4D3F3B0AA4BA89B4BCA5B&CFID=411042726&CFTOKEN=56443613&__acm__=1392562601_cc5faafc64b5b83c3dbfaa4156c4447b)
* [Novelty Detection via Answer Updating](http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA480233)
* [Bayesian game](http://he.wikipedia.org/wiki/%D7%9E%D7%A9%D7%97%D7%A7_%D7%91%D7%99%D7%99%D7%A1%D7%99%D7%90%D7%A0%D7%99)
* [Bayesian Models](http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_intromod_a0000000340.htm)
* [חוק בייס](http://he.wikipedia.org/wiki/%D7%A0%D7%95%D7%A1%D7%97%D7%AA_%D7%91%D7%99%D7%99%D7%A1)

**נספחים:**

**KeyWordExtractor UML:**

****

**Sequence Diagram:**

****

**מודל סכמתי קלט פלט:**

DB

AI

תשובה נבחרת

חישוב לפי אלגוריתם בייסיאני, מהיא התשובה המתאימה לפי מופעים של מילות מפתח.

KeyWordExtractor

שליפת מילות מפתח

קלט

שאלה מהמשתמש

פלט

**חוק בייס:**

ה"הסתברות המותנית של מאורע A בהינתן מאורע B" היא הסיכוי להתרחשותו של A, בהנחה ש- B אכן התרחש. ההנחה מכווצת, כביכול, את מרחב המדגם, וכך אפשר לחשב הסתברות מותנית על-פי הנוסחה \ P(A|B)={\frac  {P(A\cap B)}{P(B)}}.

**חוק בייס** מאפשר לחשב הסתברות שבה מניחים התרחשות של מאורע A, בעזרת הסתברויות של A, המותנות במאורעות אחרים: אם \left\{B_{i}\right\}_{{i\in K}}חלוקה של מרחב המדגם שבה לכל החלקים הסתברות חיובית, אז לכל מאורע A בעל הסתברות חיובית מתקיים (לכל K). P(B_{k}|A)={\frac  {P(A|B_{k})P(B_{k})}{\sum _{i}P(A|B_{i})P(B_{i})}}. החוק נובע ישירות מנוסחת ההסתברות השלמה.