
הבינה המלאכותית בשירות התורה

פרק ב

בפרק הקודם דיברנו על תולדות הבינה המלאכותית [ראה גם מסגרת בסוף פרק זה], מכאן נגיע לדבר העיקרי המעניין אותנו, והוא רתימתה של הבינה המלאכותית בשירות לומדי התורה. תחום שכבר עשה כמה צעדים משמעותיים בשנים האחרונות, ועוד היד נטויה.

אחת המתנות המופלאות שדורנו זכה בהן היא המאגרים האלקטרוניים העצומים של ספרים תורניים, תחום שהחלוצה בו היא תוכנת 'פרוייקט השו"ת', שהיסודות להקמתה הונחו עוד בשנת תשכ"ג, והיא היתה חלוצה גם מבחינה עולמית בתחום של שמירה דיגיטלית של טקסט לצורך חיפוש בו. שנים רבות אחריה הופיע 'התקליטור התורני', ומאוחר יותר הופיעו התוכנות המכילות תמונות של ספרים סרוקים: בשנת תשס"א התחיל המיזם של 'בית מלא ספרים' – המוכר יותר בשמה הלועזי 'היברו בוקס', ומעט מאוחר יותר 'אוצר החכמה', 'אוצרות התורה'.

המכנה המשותף של כל התוכנות האלו, מלבד עצם העובדה שהן ספריות ענק שגודלן ממלא חדרים שלמים כשמדובר בספרים ממשיים, הוא האפשרות לחפש בתוך תוכן הספרים. בתוכנות שעיקרן טקסט מוקלד כמו פרוייקט השו"ת - החיפוש מדויק מאוד, ובתוכנות כגון אוצר החכמה שעיקרן צילום של ספרים שנסרקו [ו'קריאתן' על ידי תוכנות המפענחות טקסט מתוך צילום, בטכנולוגיית OCR], החיפוש כמובן מדויק פחות, אבל ההיקף האדיר מפצה על כך. בכל מקרה, כל התוכנות האלו אינן זכאיות לתואר 'בינה מלאכותית' משום שהמחשב אינו 'מבין' כמובן את הטקסט, גם לא ברמה 'מחשבית'. הוא בסך הכל אוגר אותו בזכרונו, וכשהמשתמש מקיש צירוף אותיות מסויים, כגון 'אברהם', המחשב ימצא עבורו את הצירוף הזה בכל המאגרים שיש לו גישה אליהם. זה כמובן חיפוש פשוט מאוד, וכדי למצוא דברים מסובכים יותר, כגון משפט שהמשתמש אינו זוכר בדיוק את צורתו אלא רק את תוכנו, נדרש המשתמש לנסח את השאלתו שהוא מפנה אל המחשב בצורה מתוחכמת, לדוגמא: הלומד רוצה לחפש את משפט הפתיחה של מסכת ברכות "מאימתי קורין את שמע בערבין", אבל הוא לא בטוח אם כתוב 'בערבין' או 'בערבית', וגם לא בטוח אם כתוב 'קורין' או 'קורים' או 'קוראים' וכדומה. אז הוא יוכל להגדיר למחשב לחפש את הצורות 'קור' ו'ערב' ולא משנה איזה אותיות באות אחר כך. או - פתרון פשוט יותר – שיכתוב רק 'מאימתי שמע', ויגדיר לתוכנה שתחפש את המילים האלו גם אם יש ביניהן רווח של שתי מילים. [על פי רוב זו ברירת המחדל במנועי החיפוש]. אבל כל זה עדיין שכלולים של החיפוש הבסיסי והפשוט: מציגים למחשב צירוף אותיות מסויים, והוא מחפש ומציג את כל מופעי צירוף-האותיות הזה שקיימים אצלו במאגר. המחשב כמובן לא יודע בשום אופן את הדבר הפשוט כל כך לכל לומד, ש'ערבית' ו'ערבין' בנדון דנן הן אותו הדבר מבחינת התוצאה המבוקשת.

אבל אנו מדברים על סוג של תוכנה שתדע את הדבר הזה! תוכנה שתבין שאין הבדל בין 'ערבית' ל'ערבין', משום לשניהן אותה המשמעות, וזאת מבלי להגדיר לתוכנה הגדרות מסובכות שגוזלות זמן מן המשתמש, ומה שחמור יותר – רוב המשתמשים שאינם מומחים מתקשים מאוד לנסח את השאלות שלהן באופן המבוקש. אני, לדוגמא, משתמש 'כבד' ויומיומי בתוכנת פרוייקט השו"ת, רכשתי בה נסיון וידע רב, ועדיין אינני יודע את רוב הפונקציות המשוכללות והמתוחכמות.

פריצת הדרך הראשונה בתחום היא של 'דיקטה', עמותה בראשותו של פרופ' ר' משה קופל הי"ו, שפיתחה תשעה כלים המשתמשים באלגוריתמים מתקדמים ביותר של מה שנקרא 'למידה עמוקה' (ראה מסגרת). המכנה המשותף של הכלים האלו הוא שהם כביכול 'מבינים' את הטקסט. רחוק מאוד מסוג ההבנה שלנו, אבל כבר ברמה שמציגה תוצאות נפלאות.

לדוגמא: במנוע חיפוש רגיל, מהסוג הקיים בכל התוכנות שנזכרו לעיל, אם אני רוצה לחפש מספר כל שהוא המוזכר בגמרא, כגון 'שלוש סעודות', אצטרך לכתוב זאת במדויק. אבל אם אני לא בטוח אם במקור שאני מחפש [או בחלק מהמקורות שאני מחפש] כתוב **שלוש** סעודות, או **שלושה** סעודות [יש הרבה כאלו] או ג' סעודות - או אפילו ג סעודות בלי הגרש [יש גם כאלו], אצטרך לערוך כמה חיפושים, או לנסח את השאילתא בצורה מתוחכמת יותר, שתכלול את כל האופציות האלו, ועדיין לא אקבל את 'תלת סעודות' שבזוהר וספרי חסידות המצטטים אותו.

לעומת זאת החיפוש במנוע של דיקטה בנוי כך, שאנו יכולים לכתוב לו פשוט 3 סעודות - כך, עם הספרה 3, והוא ימצא את כל האפשרויות של הספרה 3, ואפילו צורות כמו שְלִישׁ! לדוגמא: אם נכתוב לו ברצף את המספרים 13 14 15, הוא ימצא לנו את המספרים שלושה עשר, ארבעה עשר, חמשה עשר, וגם תליסר, ארביסר וחמיסר - בתנאי שהם מופיעים בסמיכות באותו קטע גמרא (ראה תמונה).

אין זה עובד

15 14 13

מציג 1 - 5 מתוך 5 תוצאות

בבלי ומושג / סדר מועד / מסכת פסחים / פרק ה (חמדי נחש) / משנה בדף נהא / סוגיא בדף נהא

וסבר רבי יהודה **ארבעה עשר** מותר בעשיית מלאכה והתניא רבי יהודה אומר המנכש **בשלושה עשר** ונעקרה בידו שותלה במקום הטיט ואין שותלה במקום הגריד
בשלושה עשר אין **בארבעה עשר** לא מכדי שמעינן ליה לרבי יהודה דאמר כל הרכבה שאינה קולטת לשלושה ימים שוב אינה קולטת ואי סלקא דעתך **ארבעה עשר** מותר בעשיית מלאכה למה לי **שלושה עשר** והאיכא **ארביסר וחמיסר** ומקצת שיתסר אמר רבא בגליל שנו
והאיכא ליליא אמר רב ששת כבית שמאי רב אשי אמר לעולם כבית הלל לפי שאין דרכן של בני אדם לנכש בלילה
רבינא אמר לעולם ביהודה ובהרשה חד מקצת היום ככולו אמרין תרי מקצת היום ככולו לא אמרין:

בבלי ומושג / סדר מועד / מסכת יומא / פרק ב (בראשונה) / משנה בדף כג / סוגיא בדף כג

אם כן נימא קרא ובער ובערו אי נמי וערך וערכו מאי ובער וערכו שמע מינה כדקא אמרין
תני רבי חייא פייס פעמים **שלושה עשר** פעמים **ארבעה עשר** פעמים **חמשה עשר** פעמים ששה עשר והתניא שבעה עשר
ההיא דלא כרבי אליעזר בן יעקב אלא כרבי יהודה

בבלי ומושג / סדר מועד / מסכת חגיגה / פרק ב (סדר חגיגת כבוד) / משנה מדף טא עד דף יב / סוגיא בדף יב

אימת אילימא בני **חמיסר** וקא קרו ליה **בארביסר** ומי שרי
והכתיב במגילת תענית יום **ארבעה עשר** בו ויום **חמשה עשר** בו יומי פוריא אינון דלא למיספד בהון ואמר רבא לא נצרכא אלא לאסור את של זה בזה ואת של זה בזה
ואלא בני **ארביסר** וקא קרי ליה **בתליסר** יום ניקנור הוא ואלא בני **ארביסר** וקא קרי ליה בתריסר יום טוריינוס הוא

אם נחפש את הצירוף "פחות 8" נקבל גם "פחות משמונה" וגם "פחות משמינית"! הוי אומר: התוכנה לא מחפשת כאן את הצורה, אלא את הערך המספרי, והיא יודעת לקשור את המספרים לכל הצורות המילוליות שלהם. לא זו בלבד - אלא שגם אם אנסה 'להטעות' את התוכנה ולכתוב לה למשל **שנתרבה 8**, היא תמצא קטעים רחוקים, אבל את "שנתרבה שְמִנָּה" (סוטה יד:) היא לא תביא לי בשום אופן - כאן אצטרך לכתוב במפורש "שנתרבה שמנה", משום שהמילה הזאת לא שייכת לערך המספרי 8!

אין זה עובד

"פחות 8"

מציג 1 - 2 מתוך 2 תוצאות

בבלי ומושג / סדר קדשים / מסכת ערכין / פרק ב (אין מערכין) / משנה מדף טא עד דף יא / סוגיא בדף יב

אין פוחתין מארבעה חדשים המעוברים בשנה ולא נראה יתר על שמונה שתי הלחם אין נאכלין פחות משנים ולא יתר על שלשה לחם הפנים אין נאכל פחות מתשעה ולא יתר על אחד עשר
אין קטן נימול **פחות משמונה** ולא יתר על שנים עשר:

בבלי ומושג / סדר ניקין / מסכת בבא בתרא / פרק ד (המזכר את הבית) / משנה מדף סא עד דף סא / סוגיא בדף סא

פשיטא אמר יחלוק פלוני בנכסי פלגא תנו חלק לפלוני בנכסי מאי
אמר רבינא בר קיסי תא שמע דתניא האומר תנו חלק לפלוני בבור סומכוס אומר אין פחות מרביע לחבית אין **פחות משמינית** לקדרה אין פחות משנים עשר לטפיח אין פחות משישה עשר

נדרבה 8 |

מצג 1 - 2 מתוך 2 תוצאות

מצג לפי רלוונטיות

בבלי ומשנה / סדר חזקין / מסכת סנהדרין / פרק ב (כהן גדול) / משנה מדף כמא עד דף כאב / סוגיא בדף כאב

מתקין לה רבינא אימור כהנה תרתי סרי וכהנה עשרין וארבע
 תניא נמי הכי לא ירבה לו נשים יותר מעשרים וארבע למאן דדריש וזו ארבעים ושמנה הוה תניא נמי הכי
 לא ירבה לו נשים יותר מארבעים ושמנה
 ותנא דידן מאי טעמיה אמר רב כהנא מקיש כהנה בתרא לכהנה קמא מה כהנה קמא שית אף כהנה
 בתרא שית

בבלי ומשנה / סדר קדשים / מסכת מנחות / פרק א (הקומץ) / משנה מדף כמא עד דף לתא / סוגיא בדף לא

סימן סגלם תנו רבנן עושה אדם יריעה מבית שלש דפין ועד בית שמנה דפין פחות מיכן ויותר על כן לא
 יעשה
 ולא ירבה בדפין מפני שנראה אגרת ולא ימעט בדפין מפני שענינו משוטטות אלא כגון למשפחותיכם
 למשפחותיכם למשפחותיכם שלש פעמים

הרעיון שעומד מאחורי מנוע־חיפוש זה הוא שהמידע שהוא מכיל - **מתוּיג**, כלומר: מאחורי המילים שהוא מכיל, יש המון מידע באשר למורפולוגיה של המילה, כגון אם היא זכר / נקבה, פועל, שם עצם, תואר וכדומה. המידע הזה מסייע לו להבין' כביכול את הטקסט וליצור קשרים בין מילים, וכך - אם נכתוב למשל את המספר 127, יודע מנוע החיפוש בתנ"ך - שהוא המשולל ביותר - למצוא את כל הפסוקים שהמספר הזה מופיע בהם, לא משנה הסדר, וגם לא ברציפות, כגון "וַיְהִי חַי שָׁנָה וְעֶשְׂרִים שָׁנָה וְשִׁבְעִים שָׁנִים" (בראשית כג, א) וגם את "שִׁבְעֵי עֶשְׂרִים וּמֵאָה מְדִינָה" שבמגילת אסתר (ראה תמונה).

127 |

מצג 1 - 4 מתוך 4 תוצאות

מצג לפי רלוונטיות

בראשית כג, א

וַיְהִי חַי שָׁנָה וְעֶשְׂרִים שָׁנָה וְשִׁבְעִים שָׁנִים חַי שָׁנָה:

אסתר א, א

וַיְהִי בַיּוֹם אֲחֻשְׁרוּשׁ הוּא אֲחֻשְׁרוּשׁ הַמֶּלֶךְ מֵהַדּוּ וְעַד־כּוֹשׁ שִׁבְעֵי עֶשְׂרִים וּמֵאָה מְדִינָה:

אסתר ט, ט

וַיִּשְׁלַח קִפְרִים אֶל־כָּל־הַיְּהוּדִים אֶל־שִׁבְעֵי עֶשְׂרִים וּמֵאָה מְדִינָה מִלְּכוּת אֲחֻשְׁרוּשׁ דְּבָרֵי שְׁלוֹם וְאַמֶּת:

אסתר ח, ט

וַיִּקְרָאוּ סַפְרֵי־הַמֶּלֶךְ בְּעַת־הַיּוֹם הַשְּׁלִישִׁי הוּא־חֹדֶשׁ סִינּוֹן בְּשָׁלוֹשׁ עֶשְׂרִים בּוֹ וַיִּכְתֹּב כָּל־אֲשֶׁר־צִוָּה מִדְּבַרֵי אֶל־הַיְּהוּדִים וְאֵל הָאֲחֻשְׁרוּפִּיִּם וְהַפְּחוּתִים וְשָׂרֵי הַמְּדִינֹת אֲשֶׁר מֵהַדּוּ וְעַד־כּוֹשׁ שִׁבְעֵי עֶשְׂרִים וּמֵאָה מְדִינָה מִדִּינָה וּמִדִּינָה כִּכְתָּבָה וְעַם וְעַם כָּלִשְׁנוֹ וְאֶל־הַיְּהוּדִים כִּכְתָּבָם וְכָל־שׁוֹנֵם:

אם הזכרנו את מנוע החיפוש בתנ"ך, נציין כמה מיכולותיו המפליאות: סיפר לי ידיד שעייין בקינה 'שומרון קול תיתן', ונתקל שם במשפט "כי אתה סברם ואתה אָלִוְתָם" הוא רצה לחפש את המקור בתנ"ך למילה 'סבר', וחיפש לתומו את השורש 'סבר' בסמ"ך, כפי שהוא בקינה. דא עקא, במקרא כל המופעים של השורש הזה באים בשי"ן שמאלית, אך מנוע החיפוש של דיקטה יודע זאת ומביא את כל התוצאות של השורש סבר, כגון "אֵלָיו יִשְׁבְּרוּ" (תהלים קמה, טו), "וְאֵל תְּבַיְשֵׁנִי מִשְׁבְּרִי" (שם קיט, קטז), וכיוצא בזה.

בכותבנו את המילה 'דוד' - יתכן שהכוונה למלך ישראל, ויתכן שהכוונה לדוד שבשיר השירים, או לדוד, ובכן: חיפוש המילה הזאת נותן לנו 984 תוצאות, ואז יש כלי המאפשר לנו לסנן את הרצוי לנו. בשורה 'משמעויות ומילים נרדפות' אנו יכולים לבחור אם רצוננו בדוד המלך, בדוד, או בדוד או בדודאים! מה שמדהים במיוחד הוא - שאם אני אבחר, למשל, רק באופציה של 'דודאים' (צורת היחיד היא דוּדִי), תדע התוכנה להביא לי אך ורק את 'דודאים', ולא את 'דוּדָאִי תְּאֲנִים' (ירמיהו כד, א) משום שהיא לא שייכת לכאן! [ראה תמונה. וכן להיפך, אם אבחר רק ב'דוד' - אקבל את דודאי התאנים, אך לא את דודאים]. רמה כזאת של 'הבנת הטקסט' לא קיימת נכון לזמננו בשום מנוע חיפוש בעברית בעולם, וכמדומני שגם לא בשפות אחרות.

מצג 4 - 1 מתוך 4 תוצאות

מצג לפי רלוונטיות

שיר השירים ג יד
הדודאים נתנו ריח וצל פתחינו כל מגדים חדשים גם ישינים דודי צפנתי לה:

בראשית לו טו
 ותאמר לה המעט קחתה את אישי ולקחת גם את **דודאי** בני ותאמר רחל לכן ישכב עמך הלילה תחת **דודאי** בנה:

בראשית לו יד
 וילך ראובן בימי קציר חטים וימצא **דודאים** בשדה ויבא אתם אל לאה אמנו ותאמר רחל אל לאה תנינא לי **מדודאי** בנה:

בראשית לו טו
 ויבא יעקב מן השדה בערב ותצא לאה לקראתו ותאמר אלי תבוא כי שכר שכרתיך **בדודאי** בני וישכב עמה בלילה הוא:

ניחול תוצאות

משמעויות ומילים נרדפות

בחר הכל

דוד (Proper noun) David (919) ☐

דוד (Noun) beloved one (54) ☐

כלל מילים נרדפות

דוד (Noun) cooking pot (7) ☐

כלל מילים נרדפות

דודי (Noun) mandrake (4) ☒

דודה (Noun) father's sister (3) ☐

מאותה סיבה – 'הבנת' המילה, תציע לנו התוכנה גם מילים נרדפות. לדוגמא: עבור מילת החיפוש אריה, תשאל אותנו התוכנה [בצד ימין, בחלונית 'משמעויות ומילים נרדפות'] אם נרצה לבחור גם במפועים של 'יש', לביא, לבא, כפיר ושחל. אם נחפש 'כסא' – תברר אצלנו התוכנה אם אנו רוצים את כסא או קסא. לגבי 'קסא', מלבד זה שהיא יודעת גם לכלול את "שש מעלות לקסא" (מ"א, י, ט) היא תציע לנו לכלול גם את 'מושב' בתוצאות החיפוש. ולגבי קסא שפירושו ירח היא תציע לנו לכלול גם את 'לקנה'! הכלי של מילים נרדפות הוא רב עוצמה, משום שלפעמים איננו זוכרים אלא את העובדה שבפסוק מדובר על ענין שמחה, ואיננו זוכרים אם הלשון הוא שמחה או ששון או גילה או רינה. מספיק שנכתוב את אחת המילים האלו, והתוכנה תציע לנו את הנרדפים.

חיפוש בתנ"ך

בסא

מצג 5 - 1 מתוך 5 תוצאות

מצג לפי רלוונטיות

תהלים פא ד
 תקעו בחדש שופר **בכסה** ליום חגגו:

משלי ג כ
 צורו-הכסף לקח בידו ליום **הכסא** יבא ביתו:

ישעיה כד כג
 וחפרה **הלבנה** ובושה החמה כי מלך יהנה צבאות בהר ציון ובירושלם ונגד זקניו כבוד: פ

שיר השירים ג י
 מייזאת הנשקפה כמו שחר יפה **כלבנה** ברה כחמה אימה כנדגלות: ס

ישעיה לו ט
 והיה אור-**הלבנה** כאור החמה ואור החמה יהיה שבעתים כאור שבעת הימים ביום חבש יהנה את שחר עמו ומחץ מכתו ירפא: ס

ניחול תוצאות

משמעויות ומילים נרדפות

בחר הכל

בסא (Noun) seat (128) ☐

כלל מילים נרדפות

מושב ☐

בסא (Noun) full moon (5) ☒

כלל מילים נרדפות

לקנה ☐

הטיות

25 הטיות נבחרו

חיפוש בתנ"ך מבלי לדעת לשון מדויקת

התנ"ך בימינו הוא אולי הטקסט שיש לו הכי הרבה אפשרויות חיפוש. עשרות תוכנות מציעות חיפוש בתנ"ך. ככלות הכל, במושגי מחשבים הטקסט של התנ"ך הוא מעט מזער. אולם מי שמחפש פסוק מסויים ואינו זוכר בדיוק כיצד הוא כתוב, עלול לעמוד בפני תסכולים לא מעטים. שכן הכתיב המקראי מכיל חריגים רבים מאוד, כמו המילה 'חטאת' שיכולה להיות כתובה 'חטת', ראשית - 'רשית', וכדומה. בעיה מיוחדת היא שלפעמים המחפש מעוניין לדעת את כל המופעים של מילה מסויימת, אבל הוא לא יכול לסמוך על מנוע חיפוש - גם המשוכללים ביותר, משום שיתכן שכמה מופעים נמלטו מרשת החיפוש, כיון שהיו כתובים באופן חריג.

במנוע החיפוש של דיקטה הושקעו מאמצים מיוחדים, כדי שהמעין יוכל להגיע לתוצאות בכל כתיב אפשרי. לדוגמא: הכתיב הרגיל של השם המיוחד, דהיינו ה' - יביא לכל התוצאות, בין אם כתוב בפסוק שם אדנות, הויה או אלוקות. כנ"ל אם המחפש יכתוב 'אלוקים' בק' ובכתיב מלא, הוא עדיין יקבל את כל הצורות של השם הזה בה' וב כתיב חסר. וכיוצא בזה על זו הדרך. הרעיון הוא שגם בלי הוראות מדוייקות, מה שהמחפש יכתוב - הוא יצליח להגיע לכל התוצאות שהוא רוצה להגיע אליהן.

נציג כאן טבלה עם מספר שאילתות והתוצאות שהתוכנה מפיקה בעמודה הימנית כתוב הטקסט שנכתב בשאילתה, ובעמודה השמאלית - התוצאה:

חיפוש	תוצאה
סוכו	כִּי יִצְפְּנִי בְּסֻכָּה, כְּאַרְיֵה בְּסֻכָּה, וַיִּחְמַס כְּגַן שָׁכֹנִים
מי כמוך	מִי כְמוֹכָה בָּאֵלֶם ה' מִי כְמוֹכָה נֶאֱדָר בְּקִדְשׁ (שמות טו, יא)
רציתי	וְרָצָאתִי אֲתֶכֶם (יחזקאל מג, כז)
חמושים	וְחֻמָּשִׁים עָלוּ (שמות יג, יח)
יעקב	"אֵהְיִי יַעֲקֹב" (ירמיהו ל, יח)
הוא יורה	גַּם הוּא יוֹרָא (משלי יא, כה)
יובל היא	יֹבֵל הוּא תְהִיָּה לָכֶם (ויקרא כה, י)
תלבשנה	תִּלְבָּשׁוּן בְּנוֹת הַמֶּלֶךְ (ש"ב יג, יח)
חיטים	וְאַתָּה קָח לָךְ חֹטִין וְשֹׁעֲרִים (יחזקאל ד, ט)
סחט	וְאָקַח אֶת הָעֲנָבִים וְאֶשְׁחַט אֹתָם אֶל כּוֹס פְּרָעָה (בראשית מ, יא)
ברא תנינים	וַיִּבְרָא אֱלֹהִים אֶת הַתְּנִינִם הַגְּדֹלִים (בראשית א, כא)
מילים אמר	רֹאשׁ מַלְיָן אָמַר (דניאל ז, א)
נאום ה' אלקים	נֹאם אֲ-דְנִי יְדֹד (ירמיה ב, יט ועוד הרבה)
ויאמר ל משה	וַיֹּאמֶר לָהֶם מֹשֶׁה (שמות יז, ב); וַיֹּאמֶר לוֹ מֹשֶׁה
יום השישי ויכולו	וַיְהִי עֶרֶב וַיְהִי בֹקֶר יוֹם הַשְּׁשִׁי (בראשית א, לא), וַיְכֹלּוּ הַשָּׁמַיִם וְהָאָרֶץ (בראשית ב, א) [התוכנה מגשרת על שני פסוקים ושני פרקים]
300 50 30	שָׁלֹשׁ מֵאוֹת אִמָּה אֶרֶךְ הַתֵּבָה חֻמָּשִׁים אִמָּה רְחֵבָה וּשְׁלֹשִׁים אִמָּה קוֹמָתָה (בראשית ו, טו)

כתיב מלא וחסר

למעשה הענין של כתיב מלא וחסר מקבל מענה גם בתוכנות כמו התקליטור התורני או פרווייקט השו"ת ותוכנות נוספות. יש שם אופציה לכתוב את המילה עם כתיב מלא, ולהורות למנוע החיפוש לחפש את התוצאות גם כשהכתיב חסר, ולכן נוכל לכתוב 'יעקוב' ולקבל גם 'יעקב' [להיפך אי אפשר: כלומר, 'יעקב' לא יתן את 'יעקוב', בתקליטור התורני זה עובד הפוך: כותבים את הצורה החסרה ומורים לתוכנה לחפש גם צורות מלאות].

אבל גם המענה הזה עדיין אינו מתקרב לשלל האפשרויות שהדגמנו לעיל, ונוסיף עוד נקודה אחת: במנועי החיפוש הרגילים - ברגע שאתה מורה למנוע לחפש גם צורות השונות מהכתיב שהקלדת, אתה מסתכן בתוצאות רבות שאינן רלוונטיות. לדוגמא: אנו רוצים לחפש את המילה 'חֻמָּשִׁים' בתנ"ך, ומודעים לאפשרות שיתכן שהיא תהיה כתובה בכתיב חסר. לכן נסמן במנוע החיפוש של פרווייקט השו"ת את האופציה של חיפוש בכתיב חסר. ומה שיקרה הוא שאנו נקבל 159 תוצאות שרובן ככולן היא המילה 'חֻמָּשִׁים' - גם היא באותו הכתיב בדיוק. כדי להתגבר על זה נצטרך לסמן לתוכנה לחפש לפי ניקוד - יש את האופציה הזאת בפרוייקט השו"ת, אבל היא כבר אופציה "מטיבי לכת", וגם היא עובדת רק בטקסט התנ"ך ולא יותר.

לעומת זאת במנוע החיפוש של דיקטה מספיק שנכתוב 'חמושים' בלי שום הוראה נוספת ונקבל את כל ארבע התוצאות שיש במקרא, בלי תוצאות לא רלוונטיות שנצטרך לחפש בתוכן, וכולל גם את הצורות 'החמושים' 'וחמושים'.

המשך במאמר הבא בעז"ה

בינה מלאכותית - רקע

למידה עמוקה

הטכנולוגיה שנקראת 'למידה עמוקה', ובעבר 'רשת נוירונים מלאכותית' היא היום הדבר ה'חם' ביותר בתחום הבינה המלאכותית, ופריצות הדרך המשמעותיות בשנים האחרונות, כגון תוכנות זיהוי פנים, תרגומי שפה ומכוניות אוטונומיות מבוססות עליה. הנושא הזה מורכב וגדול מדי בשביל יריעה זו, אך נציג כמה דברים על קצה המזלג, הלקוחים ממאמרו המרתק של רן לוי, 'למידה עמוקה – המחשב והמוח'.

כדי לנסות ולבנות 'מכונה חושבת', פנו החוקרים אל הביולוגיה ואל מדעי המוח, כדרך רוב המכונות שנבנו על סמך חיקוי של דברים הקיימים בבריאה. התקדמויות בתחום הנוירולוגיה והפיזיולוגיה של המוח סיפקו להם השראה למכביר, במיוחד לאור העובדה שלאורך המאה העשרים נחשפו מנגנונים ביולוגיים רבים שעומדים בבסיס פעילותו של המוח. עם הזמן החלו המדענים להבין את חשיבות הנוירונים, תאי העצב שמהם עשוי המוח. הנוירונים הם תאים זעירים בעלי זרועות ארוכות שמתחברים זה לזה באמצעות חללים הנקראים 'סינפסות' ומעבירים ביניהם מידע באמצעות זרמים חשמליים ותגובות כימיות. לכל נוירון יש מספר כניסות, והוא מקבל 'פולסים' מנוירונים אחרים. בתגובה לפולסים אלה מפיק הנוירון פולס חשמלי משלו ביציאה. טריליוני החיבורים שבין מיליארדי תאי העצב והפולסים שהם מעבירים ביניהם מהווים, בדרך נסתרת כלשהי, את התשתית לכל הפעילות המוחית שלנו.

בשנת תש"ט גילה חוקר בשם דונלד הֶבְּ עיקרון חשוב מאוד בדרך הפעולה המוחית, הדרך שבה מתחוללת למידה. הב גילה שאם נוירון א' ונוירון ב' מחוברים ביניהם, ונוירון א' יורה פולסים חשמליים לאורך זמן – נוירון ב' יתחיל לירות פולסים ביעילות גבוהה יותר. דהיינו – מתרחש כאן תהליך שבו ירי הפולסים הבלתי פוסק של נוירון א' יגרום לכך שנוירון ב' ילמד כי המידע שמגיע מנוירון א' הוא מידע חשוב ויש להגיב עליו באמצעות ירי של פולס משלך בתגובה.

התובנה הזו העניקה השראה לפסיכולוג בשם פרנק רוזנבלט, מומחה לקוגניציה מאוניברסיטת קורנל שבארצות הברית. בשנת תשי"ח הגה רוזנבלט רכיב חשמלי מסוג חדש שכונה 'פֶּרְצֶפְטוֹן' מלשון 'Perception', (פֶּרְצֶפְטִין - תפיסה).

הפרצפטור של רוזנבלט היה מעין 'נוירון מלאכותי', מודל מופשט של הנוירון האנושי. היו לו מספר כניסות שקיבלו ערכים בינאריים – 0 או 1 – ויציאה אחת שיכולה גם היא להפיק 0 או 1. אם מספיק כניסות לפרצפטור קיבלו את הערך '1' – הגיב המכשיר גם הוא ביצירת 1. הפרט המעניין הוא שרוזנבלט מצא דרך לדמות את תהליך הלמידה שתיאר הנוירולוג הֶבְּ; באמצעות נוסחה שהוא בנה, הוא הצליח 'ללמד' את הפרצפטור שפולסים המגיעים מכניסה פלונית חשובים יותר ויש להגיב עליהם יותר מאשר אלו הנכנסים בכניסה אחרת, ביכולתו היה לשלוט במשקל שינתן לכל כניסה, וכך זה היה חיקוי פרימיטיבי לעוצמת הקשרים שנוצרים בין נוירונים ביולוגיים במוח. משהצליח לבנות מודל ששולט על שטף המידע הנכנס, הקים במעבדתו מכשיר שהכיל כמה פרצפטורים שכאלה מחוברים זה לזה במעין רשת של 'נוירונים מלאכותיים', וחיבר את הכניסות שלהם לארבע מאות קולטני אור. רוזנבלט הציב מול קולטני האור אותיות, ספרות וצורות גאומטריות – ובאמצעות כיוון עדין של העוצמה בכניסות הפרצפטורים הצליח רוזנבלט 'ללמד' אותם לזהות את הצורות ובתגובה להפיק במוצאם אותות שאומרים – 'זו האות א' או 'זה מרובע'.

כיצד עשה זאת? באמצעות חיזוק או החלשת הקשרים שבין הפרצפטורים: בכל פעם שהפרצפטורים לא זיהו נכון צורה כלשהי, הוא שינה מעט את ה'חשיבות' של כניסה מסוימת עד שכל פרצפטור 'למד' שצירוף מסוים של כניסות הוא א', צירוף אחר הוא ב' וכדומה. המשחק עם הכניסות הוא דרך לתקן את השגיאות של המערכת, לומר לה – "מה שעשית עכשיו היה לא נכון. הנה הדרך הנכונה לעשות זאת." את ה'משחק' בין הכניסות קבעה נוסחה שהוא חישב.

רוזנבלט הצליח ללמד את המכשיר - לא באמצעות כללים מוכתבים, כמו שעושים תמיד, אלא באמצעות הצגה של מגוון דוגמות לצורה מסוימת, וחיזוק או החלשת הקשרים בין הפרצפטורנים עד שנמצא הצירוף הנכון שמאפשר למערכת כולה לזהות את הצורה הזאת. כלומר: אם בדרך הרגילה נותנים למחשב פשוט איזה כלל, למשל - "אם לצורה יש ארבע צלעות, אז היא מרובע" - רוזנבלט סיפק למכשיר שלו דוגמאות של מרובעים ובאף שלב לא הגדיר לפרצפטורנים כמה צלעות יש למרובע. הוא רק החליש והגביר את העוצמה של כניסות שונות, עד שהמערכת 'למדה' להכיר מרובעים. גישה זו של לימוד מתוך דוגמאות, קרובה מאד לדרך שבה בני אדם רוכשים כישורים רבים: ילדים בגיל רך לומדים כך בדרך כלל את רוב המידע שהם רוכשים, מזיהוי צורות ועד שפה. מראים להם דברים, הם טועים, מתקנים אותם, הם טועים שוב ולומדים שוב עד שהם רוכשים בעצמם את התחושה ש'מרובע' זה הדבר הזה, גם בלי הגדרה מדויקת.

כשהתפרסמה העבודה של רוזנבלט היה העולם כמרקחה, והעיתונים התחרו זה בזה בהפלגות ותיאורים של מה שתצליח 'בינה מלאכותית' לעשות. ה"ניו-יורק טיימס" סיפר לקוראיו על המחקר של רוזנבלט, שנערך במימון הצי של צבא ארה"ב, במילים הבאות: "מכשיר חדש של הצי לומד באמצעות עשייה! הצי חשף היום את אב הטיפוס הראשוני של מחשב אלקטרוני שעל פי הציפיות יהיו מסוגל ללכת, לדבר, לראות, לכתוב, לשכפל את עצמו ולהיות מודע לעצם קיומו", מה שנקרא "בלי עין הרע". רוזנבלט עצמו אמר באותו הראיון כי הוא מאמין שבעתיד יוכלו מכונות אלה לטוס במרחבי מערכת השמש ולבצע פעולות שונות בכוחות עצמן. חוקרים אחרים שיערו שהמחשבים בעלי הבינה המלאכותית יוכלו בעוד מספר שנים לתרגם באופן מיידי מרוסית לאנגלית ולהפך, ובכך יתרמו לסיום המלחמה הקרה. ההתלהבות היתה בשיאה, ואלפי מחקרים ומאמרים אקדמיים הוקדשו לפרצפטורנים ולפוטנציאל הטמון בהם.

אולם כדרך תגליות רבות הסתיימה ההתלהבות בקול דממה דקה; מרווין מינסקי, מתמטיקאי אמריקני שנחשב לאחד החלוצים הגדולים בתחום הבינה המלאכותית, טען בנחרצות שהמכשיר של רוזנבלט לא יכול להגיע לכדי בינה מלאכותית. ועשר שנים לאחר שהדגים רוזנבלט לראשונה את אב-הטיפוס של הפרצפטורן, הוציא מינסקי יחד עם מתמטיקאי נוסף בשם סימור פאפרט, ספר בשם 'פרצפטורנס'. בספר ניתחו ניתוח תאורטי מעמיק את הרכיב שהמציא רוזנבלט - והגיעו למסקנה שיהיה זה בלתי אפשרי לפתח בעזרתו מערכות בעלות בינה מלאכותית. משום שהמודל שלו פשטני לאין ערוך, וכדי להגיע לתוצאות מסובכות מעט יותר מביצועי הפרצפטורן צריך מכשיר בעל 'שכבות רבות', ופיתוח מכשיר כזה הוא מן הנמנע. לדעתם של מינסקי ופאפרט נזירות מלאכותיים הם מבוי סתום, אין טעם להמשיך ולעסוק בהם, ועדיף לחוקרים שינסו לממש את חלום הבינה המלאכותית בדרכים אחרות.

הספר הזה, יחד עם העובדה שגם בפועל לא הצליחו להתקדם הלאה, מעבר למודל הראשוני, היה מכת מוות לכל רעיון הנזירות המלאכותיים, ומשנות השבעים ואילך נזנח התחום לגמרי, עד כדי ששום מגזין מדעי לא הסכים לפרסם מאמרים על רשתות נזירות, שלא לדבר על מענקים למדענים שרצו להמשיך לחקור את הנושא. מדען בשם פול וורבוס הצליח למצוא בשנת תשל"ד פתרון לבעיה שהצביעו עליה מינסקי ושותפו, אולם קולו היה כקול קורא במדבר, אף אחד לא הסכים להקשיב לו. רק כעשר שנים מאוחר יותר התעורר העניין מחדש, מכיון מפתיע ביותר: שני פסיכולוגים, דיוויד רומלהארט וג'יימס מק'לילנד לקחו חלק בויכוח שהתנהל בקהילה האקדמית בשאלת יכולת הדיבור האנושי, שכבר עסקנו בה פעם. יכולת הדיבור מיוחדת למין האנושי, והיא דבר מופלא שעד היום לא ברור כיצד הוא עובד.

שני הפסיכולוגים האלו רצו לחזק תפיסה מסוימת בתחום הלמידה הלשונית, ולשם כך פנו אל רשת נזירות מלאכותית ו'לימדו' אותה כיצד להטות פעלים באנגלית. אנו נביא דוגמה מעברית, אבל הרעיון הוא זהה.

אחד הדברים המעניינים ביותר בתחום למידת שפה אצל ילדים קטנים הם הטעויות המיוחדות שהם עושים בשלבים מתקדמים יותר בלימוד השפה. בתחילה ילד מחקה את המילים שהוא שמע, ככל שהוא מסוגל. השיבושים שלו הם בדרך כלל חוסר יכולת להגות עיצורים מסוימים, או שהוא לא שמע את המילה באופן ברור. אך בשלב מתקדם יותר הוא מתחיל 'להמציא' מילים חדשות. לדוגמה: ילדים קטנים אומרים 'אישות' כצורת רבים של אשה, ולא 'נשים'. הם אומרים 'לנסוע' או 'נסע' במקום לטוע, יסע. וכן **שתי** את המים,

ראיתי מכונית, **תלתי** כביסה - במקום **שתייתי**, **ראיתי**, **תליתי**, כל זאת – אף שהם לא שמעו בסביבתם מעולם את הצורות האלו.

כשמתבוננים בטעויות כגון אלו, רואים דבר מדהים; כל המילים האלו הן חריגות, ואינן מתאימות לצורה הרגילה. צורת הרבים של 'אשה' באמת מתבקשת להיות 'אישות' [ויש אפילו פעם אחת כזו בתנ"ך בספר יחזקאל (כג, מד)]. גם המלה **שתייתי** היא חריגה. בצורה הרגילה אומרים **לקחתי**, **אמרתתי**. אבל בשרשים המסתיימים בה"א כמו שתה, ראה, קנה - משתנה הניקוד, ואומרים אותם עם חיריק בע' הפועל. יש לחריגות זאת הסבר לפי כללי הדקדוק, אבל התינוק בן השנתיים מקיש מהפעלים האחרים שהוא מכיר, ואומר **שתייתי** כמו **לקחתי**. כנ"ל ביחס למלה 'לנסוע'; מהמלה **שבר** אומרים **לשבור**, וכך **מנסע** היו צריכים לומר **לנסוע**, אלא שלשרשים המתחילים בנו"ן יש תכונה מיוחדת בדקדוק הלשון העברית, והנו"ן 'נושרת' לפעמים. אבל הילד מבין שכשם שאומרים 'לשבור' ולא 'לבור' כך צריכים לומר 'לנסוע'.

בשלב מאוחר יותר, לומד הילד להטמיע את כל החריגויות האלו, למעט מקרים מסובכים יותר שנשארים גם בשפתם של מבוגרים, כגון **שבור** בב' דגושה או **לפזק** בפ' רפה, והרבה כיוצא בזה.

באנגלית, כמו בכל שפה בעולם יש גם ערימות של חריגים, ובאנגלית במיוחד. אחת הדוגמאות הנפוצות היא תופעת הטיית הפעלים בעבר: ברגיל – כדי לקבוע זמן עבר לפועל מוסיפים לו את האותיות ED, לדוגמא, המילה work (וֹרְק) הופכת בעבר ל-worked. אבל יש כמות נכבדה של פעלים חריגים שצורת העבר שלהם נעשית באופנים אחרים, והם כאב ראש רציני לכל מי שלומד אנגלית. בדרך כלל ילדים קטנים בשלב המתקדם יותר 'מלבישים' את הסיומת ED הרגילה גם לפעלים החריגים, עד שהם מתבגרים ולומדים לתקן את טעויותיהם.

למרבה ההפתעה, רשת הניירונים המלאכותיים של רומלהארט ומק'קלילנד **עשתה אותו מסלול הלימוד כמו זה של הילדים!** בתחילת תהליך הלימוד המערכת ניבאה נכון את צורת העבר של פעלים – אבל אז, ככל שהזינו לתוכה יותר ויותר דוגמאות, רשת הניירונים זיהתה את החוקיות של הוספת ed לפעלים – ואז, ממש כמו אצל ילדים, היא החלה **לטעות** ולהוסיף ed אפה שלא צריך. רק כשהזינו החוקרים לתוכה עוד ועוד דוגמאות של הטיית פעלים למדה המערכת מתי צריך להוסיף ed ומתי לא – ויכולת הניבוי שלה השתפרה בהתאם.

התופעה הזאת הדגימה – ראשית כל – שרשת ניירונים יכולה ללמוד מאפיין של שפה אנושית **אך ורק מתוך דוגמאות** ולא בעזרת ניסוח פורמלי של כללים. לא זו בלבד, אלא שרשת הניירונים המלאכותית הפגינה תהליך לימוד 'אנושי' להפליא שחיקה לא רק את ההצלחות האנושיות, אלא גם את אותו סוג של טעויות שעושים בני אדם! ניסוי זה נחשב לתפנית גדולה, והצית מחדש התלהבות עצומה לתחום רשתות הניירונים. בין השאר, השם שלו שונה 'למידה עמוקה' במקום 'רשתות ניירונים' כדי להיפטר מהסטיגמה שדבקה לתחום מאז הספר של מינסקי ופאפרט. קפיצות דרך גדולות התרחשו כשהצליחו לפתח מחשבים מהירים יותר ויותר, במיוחד לאחר שהופיעו בעולם 'מעבדים גרפיים', טכניקה של אלפי מעבדים זעירים שכל אחד מהם אינו חזק כשלעצמו, אבל העובדה שכולם עובדים ביחד מקפיצה את כוח החישוב שלהם לעוצמות אדירות. בתחום 'ראיה ממוחשבת' (זיהוי נכון של תמונות, פנים למשל), היתה קפיצה בבת אחת מיכולת זיהוי של 26% ל-76%, זה היה בשנת תשע"ב, שנה לאחר מכן כבר הגיעו לזיהוי נכון של 89%, ובשנת תשע"ה הצליחו להגיע ל-97% דיוק! זו אגב תכונה שכל אחד מאיתנו יכול לראות במצלמות דיגיטליות הקיימות בימינו במחיר שווה לכל כיס: למצלמה יש יכולת זיהוי פנים, ומיד כשהחיישן האלקטרוני 'רואה' פנים מולו, הוא מסמן אותם במין ריבוע. זו יכולת שנראית לנו כדבר של מה בכך, אולם לקחו עשרות שנים עד שהצליחו להנחיל יכולת זו למכונות. וכבר הזכרנו את המכונות האוטונומיות שכבר מתחילות לנסוע על הכבישים באירופה וארה"ב. מלבד זה יש עוד רשימה ארוכה, ותחומים חדשים מתווספים מדי יום.

הכלים של 'דיקטה' שהזכרנו בתוך המאמר, במיוחד הנקדן שעדיין לא הספקנו לדבר אודותיו, משתמשים בין השאר גם ביכולת זו של 'למידה עמוקה'. כלומר, הם לומדים מדוגמאות, וככל שהם נחשפים ליותר טקסטים – היכולת שלהם משתפרת.