

המחשב שיפענח בתוך שניות טקסטים עתיקים רבים יותר משיספיק חוקר בכל חייו

באוניברסיטת חיפה מגייסים את ההייטק כדי להזרים רוח חדשה ללימודי ההיסטוריה, המקרא והספרות. פרופ' צביקה קופליק: "הזמנים שבהם חוקר קורא ערימות של ספרים עומדים להסתיים"

עופר אדרת 17.01.2020 01:50

על צגי המחשב במעבדת מדעי הרוח הדיגיטליים באוניברסיטת חיפה הופיעו בחודשים האחרונים משפטים מוזרים. "תמהר כדה אל השוקו", כתב המחשב באחת הפעמים. "הפה ביקש לעשן", התעקש פעם אחרת והיתה גם פעם שבה קבע: "ישו ה' לשוש". מאחורי החיוכים של שני מנהלי המעבדה — האחד ד"ר לתלמוד והשני פרופ' למערכות מידע — הסתתר גם סיפוק רב. הבייבי שלהם, כלי ממוחשב העונה לשם "קראקן", החל לעשות את צעדיו הראשונים בלימוד עברית עתיקה. ממש כמו ילדים, שנפגשים לראשונה עם המקורות העבריים בבית הספר היסודי, גם קראקן זקוק לתרגול, בטרם ישחה בחומר. ה"שוקו" מהמשפט הראשון אמור היה להיות "שוקת". הפה, מהמשפט השני, ביקש "לעסוק בתורה" (ולא לעשן) ואילו ישו, ה' ישמור, כלל לא קשור למשפט השלישי, שבמקור נוסח כך: "ישוב ה' לשוש".

לפי התחזיות, בעתיד הלא רחוק, לאחר שהמחשב ישלים את חוק לימודיו, הוא יוכל לקרוא בעצמו כל טקסט עברי שיונח לפתחו, גם אם הכתב בו בלתי קריא ומשובש וגם אם תוכנו קשה לפענוח. אחרית לכך דיסציפלינה חדשה יחסית, "מדעי רוח דיגיטליים" שמה, שרותמת את המלה האחרונה בטכנולוגיה לשירותם של חוקרים בתחומים כמו היסטוריה, מקרא וספרות.

ד"ר משה לביא, בוגר 8200, מרצה לתלמוד ומדרש ואחד משני החוקרים העומדים בראש המעבדה בחיפה, הפנה השבוע לרגע את מבטו מצג המחשב, כדי להציג את מרכולתו. בהתלהבות מדבקת הוא סיפר על המהפכה הדיגיטלית, שעתידה להציל מכליה תחומי מחקר ותיקים רבים, שמיהרו להספיד בשנים האחרונות. לחלל האוויר הוא זרק ביטויים שבמבט ראשון רחוקים שנות אור מתחומי המחקר "המיושנים" שלו: "למידה עמוקה", "קריאה רחוקה", "ראייה ממוחשבת", "כריית נתונים" ו"רשתות נוירונים מלאכותיות".

על המסך הוא הציג קטע סרוק ממדרש תנחומא, אוסף מדרשי אגדה ארצישראלי משלהי העת העתיקה. כתב היד הזה קשה לקריאה, אך המחשב לא מרים ידיים. "קראקן" — פרי פיתוח של פרופ' דניאל שטוקל בן-עזרא מ"אקול פראטיק דה אוטז אטיד" בפרז — מצליח לקרוא אותו, ובהמשך מגיש אותו לחוקר על מגש של כסף, כקובץ טקסט פשוט.

בכך הוא פותח בפני החוקרים אפשרויות חדשות שמציתות את הדמיון, ובראשן חיפוש וניתוח מידע בהיקפים ובסוגי טקסטים מסוג שעד כה גם החוקר המיומן ביותר לא היה יכול לבצע בכוחות עצמו. "החזון הגדול שלנו הוא הנגשה טקסטואלית של כל כתבי היד העבריים", אומר לביא. "אנחנו נהפוך את המורשת היהודית והעברית לטקסטים זמינים לחיפוש, לימוד ומחקר, ונציל אוצר בלום של ידע ומסורות יהודיות".

כדי להבין טוב יותר על מה הוא מדבר, יש להכיר את השינויים הטכנולוגיים שעוברים בשנים האחרונות בעולם כתבי היד. בעבר טקסטים עבריים עתיקים היו זמינים בספרים בלבד. כדי לעיין בהם נדרשו החוקרים לשבת בספריות כשלצדם ערמות של ספרים עבי כרס. בשנים האחרונות מבוצעים

ברחבי העולם פרויקטים של דיגיטציה של כתבי היד האלה, במסגרתם הם נסרקים ומועלים למחשב קבוצי תמונה. השלב הבא, שבו מתמקדת כעת מעבדת מדעי הרוח הדיגיטליים, נועד להפוך את הקבצים האלה מתמונות — שרק העין האנושית יודעת לקרוא את האותיות שניבטות מבעדן — לקובצי טקסט, מהסוג שהמחשב יודע לקרוא.

טכנולוגיית "זיהוי אוטומטי של כתבי יד" (HTR) עומדת מאחורי המהפכה, שתאפשר למחשב לקרוא עשרות אלפי עמודים — דוגמת רומנים ושירה מהמאה ה-19, יומנים ומכתבים ממלחמת העולם השנייה וטקסטים פילוסופיים ודתיים עתיקים. במסגרתה, לדברי לביא, "מלמדים את המחשב לקרוא את הטקסטים באופן אוטומטי באמצעות למידה חישובית, תוך התבססות על 'אימון' שלו, כך שהמחשב רוכש גם ידע הקשרי על השפה, ומשתמש בו כדי להגיע לתוצאות טובות יותר". פרופ' צביקה קופליק, שותפו לניהול המעבדה, מוסיף ש"הזמנים שבהם חוקר מבלה בקריאת ערימות של ספרים בספריות מאובקות, עומדים להסתיים". לדבריו, "הטכנולוגיה יכולה לחסוך מהחוקרים את החיפוש הסיזיפי, ולאפשר להם לנצל באופן יעיל יותר את הזמן שלהם, כדי להפנות אותו לאפיקים חשובים יותר".

בשלב הזה, כאמור, המחשב עדיין זקוק לעזרה מהחוקרים, שמלמדים אותו לקרוא ו"להבין" את הטקסטים העבריים העתיקים שבהם הוא נתקל בפעם הראשונה. "אנחנו מציגים למחשב תמונות רבות מתוך כתבי יד, לצד התעתיק הנכון שלהן" אומר לביא. "המחשב מוצא בעצמו את הנוסחה המתמטית המובילה מהנתונים הוויזואליים לטקסט, וכך מפתח את היכולת לפענח גם את הטקסט בכתב היד שאת התעתיק שלו לא הראו לו קודם".

דרור אלוביץ, המנהל הטכנולוגי של המעבדה וסטודנט לתואר שני בהיסטוריה, מעריך כי "לא רחוק היום שלא נצטרך יותר את הגורם האנושי, והטקסטים יעברו דיגיטציה בעצמם". אלוביץ, שעבד במשך עשור בסטרטאפים ישראליים, החליט לפני כמה שנים לקחת הפסקה מהמרוץ והחל ללמוד היסטוריה באוניברסיטה. באחד השיעורים, שעסק בקיסרות הרומית, חשף אותו המרצה לעולם המתפתח של מדעי הרוח הדיגיטליים, והמחיש לו כיצד הוא יכול לשלב בין שני העולמות — זה שהותיר מאחור בחברות ההיי-טק הנוצצות, וזה שאליו הצטרף כעת, במסדרונות המיושנים של הפקולטה למדעי הרוח.

"דיגיטל יומניטיז" בלשונו, נתפס עדיין בציבור הרחב וגם באקדמיה "כתחום זר, לא מוכר ומאיים". אבל בשנים האחרונות הוא מתקרב יותר ויותר ל"מיינסטרים" של האקדמיה והופך, לדבריו, "לחלק בלתי נפרד ממדעי הרוח". כיום גם באוניברסיטאות העברית, בר אילן ובן גוריון מתנסים בתחום הזה. בחזונו של אלוביץ, כבר בשנה א' של לימודי ההיסטוריה, לצד מבוא לעת העתיקה, הסטודנטים יידרשו להכיר את הכלים החישוביים — מהתוכנות הפשוטות ועד האלגוריתמים הייעודיים, שיעשירו את המחקר ההומניסטי הוותיק.

המחשה למהפכה הדיגיטלית במדעי הרוח קיבל אלוביץ באחת מעבודות המחקר שנדרש להגיש. הוא השווה בין כמות החומר הגולמי שאסף פרופסור שחקר את אותו נושא, במשך שבועות של עבודה בארכיונים, ובין החומר שהוא עצמו הצליח להוציא בכמה דקות בסיוע המחשב מהבית. "הוא הגיע רק ל-10% ממה שאספתי בעצמי", אומר אלוביץ. בחוויה שלו יכול להתנסות כל אחד מקוראי הכתבה הזו כבר היום. הוא השתמש במאגר המידע החופשי שבאתר "עיתונות יהודית היסטורית" (JPress), שבו סרוקים 2.5 מיליון עמודי עיתונים מהמאה ה-19 ועד היום — פרויקט שייסד פרופ' ירון צור מהחוג להיסטוריה של עם ישראל באוניברסיטת תל אביב, אחד מחלוצי מדעי הרוח הדיגיטליים בארץ. "לא משנה כמה משאבים יהיו לך כחוקר בודד ולא חשוב כמה גדולה קבוצת המחקר שלה, בסופו של דבר, לא תוכל לגעת אלא בקצה קצהו של הקורפוס ולא תוכל לכבוש אלא חלקיק קטן מהעושר של כתבי היד הקיימים", מסביר אלוביץ. בזכות המחשב, החוקרים יכולים לעיין בבת אחת בכל החומר "ולהפעיל עליו כלים מחקרניים של ניתוח טקסטואלי, כדי לגבש תובנות מחקריות חדשות", כדבריו.

הפרויקט שמובילה מעבדת מדעי הרוח הדיגיטליים החיפאית נועד לתת מענה מקיף וכולל לטקסטים עבריים באשר הם. בעתיד הוא יוכל גם לשדרג מיזמים קיימים, שעשו בשנים האחרונות דיגיטציה לכתבי יד כמו הגניזה הקהירית ומגילות ים המלח, אך השתמשו בטכנולוגיה אחרת, ועדיין לא מאפשרים מיצוי של יכולות המחשב לקרוא ולעבד את הטקסט.

האדריכלית רות קפלן, דוקטורנטית בחוג להיסטוריה של עם ישראל באוניברסיטת חיפה, רתמה אף היא את מעבדת מדעי הרוח הדיגיטליים לשירות המחקר שלה, שעוסק בהיסטוריה העירונית של יהודי לודז'. כדי להכיר יותר את מושא המחקר שלה היא מנתחת, בסיוע המחשב, מפקדי אוכלוסין שבוצעו בקרב הקהילה. "מהמפקדים האלה ניתן להפיק מידע רב ערך על החיים שנעלמו לאחר עליית הנאצים לשלטון, אם רק יודעים איך לנצל נכון את היכולות של המחשב", היא אומרת.

את תוצרי המפקדים היא איתרה בארכיונים בפולין, בפורמט מיושן ובאיכות נמוכה. "כעת, לאחר שהמחשב למד לקרוא אותם, ניתן יהיה להפיק מהטקסטים האלה — באמצעות כלים ניתוחיים ממוחשבים — מידע שעד כה הסתתר בהם", היא מוסיפה.

"אנחנו לא רוצים לקבור את מדעי הרוח הקלאסיים", מדגיש לביא. "להיפך, אנחנו רוצים להציל אותם. אנחנו לא מחליפים את החוקרים של פעם, אלא משפרים את המוח האנושי באמצעות כלים ממוחשבים".