

מקהיר העתיקה לרשת האינטרנט – פרויקט הדיגיטציה של גניזת קהיר רוני שויקה, פרויקט פרידברג לחקר הגניזה

גניזת קהיר

בשם "גניזת קהיר" אנו מכנים אוסף ענק של כתבי-יד ותעודות שהתגלו בסוף המאה ה-19 בבית כנסת עתיק בקהיר. בית הכנסת נבנה בפוסטאט, היא קהיר הקדומה, במאה התשיעית לספירה, ומאז שימש את הקהילה היהודית במצרים לאורך מאות שנים. מכיוון שעל-פי ההלכה היהודית אסור לזרוק לאשפה את כתבי הקודש שבלו, הוקצה חדר בעליית הגג של בית הכנסת, סמוך לעזרת הנשים, בו גנזו ואליו הושלכו במשך מאות שנים כתבי-יד של ספרי קודש, אך גם של ספרים אחרים, ולמעשה כל פיסת נייר שהייתה כתובה באותיות עבריות. כך הצטבר בחדר זה מטמון תרבותי חסר תקדים בהיקפו ובחשיבותו, מעין "קפסולת זמן" המתעדת את החיים היהודיים על כל היבטיהם מהמאה התשיעית ועד המאה ה-19, אם כי עיקר החומר שייך למאה העשירית עד המאה ה-14. בין התחומים הרבים שהגניזה הקהירית מייצגת נמצאים בין השאר מקרא ופרשנות, ספרות תלמודית ורבנית, ספרות הגות ומדע, ליטורגיקה ושירה, חכמת הלשון וכתבים רפואיים, מאגיה, אסטרולוגיה ואסטרונומיה, קבלה ומיסטיקה, ועוד. חלק ניכר מספרות זו כתוב בערבית-יהודית, היינו בשפה ערבית הכתובה באותיות עבריות. מלבד החומר הספרותי המגוון, שבמקרים רבים נשתמר אך ורק בגניזה ולא היה ידוע לנו ממקום אחר, הגניזה שימרה עבורנו באופן חסר תקדים בתרבות האנושית גם אוצר גדול של תעודות אישיות, מכתבים, מסמכים רשמיים, המחאות ושטרי חוב, גטין וכתובות, פסקי-דין ותעודות היסטוריות, המתעדים את חיי היום-יום של קהילות היהודים במזרח בין המאה ה-10 למאה ה-14.

פיזור הגניזה

מאז גילוי הגניזה בסוף המאה ה-19 והעברת חלק ניכר ממנה לספריית אוניברסיטת קמברידג' על-ידי סלומון שכטר בשנת 1897, עוסקים החוקרים כבר למעלה ממאה שנה בלימוד, פרסום ומחקר של הקטעים שעלו ממנה. עם השנים התפזרה הגניזה לספריות ואוספים רבים, וכיום קטעי הגניזה נמצאים בקרוב לשבעים ספריות ואוספים שונים. האוסף הגדול ביותר נמצא כאמור בספריית אוניברסיטת קמברידג'. אוספים גדולים אחרים נמצאים בספריית הבודליאנה באוקספורד, בספרייה הבריטית בלונדון, בספריית חברת "כל ישראל חברים" בפריז, בספרייה הלאומית בירושלים, בספרייה הלאומית של רוסיה בסנט פטרבורג, בבודפסט, בווינה, בז'נבה, בספריית בית המדרש לרבנים בניו-יורק, בפילדלפיה, ועוד. סך הכל, מספר הדפים שעלו מהגניזה מגיע ל-320,000 דף, הרשומים תחת כרבע מיליון מספרי מדף שונים בספריות השונות.

פרויקט פרידברג לחקר הגניזה

לפני כעשר שנים ייסד דב פרידברג, נדבן ובעל הון מקנדה, את "פרויקט פרידברג לחקר הגניזה". מטרתו של הפרויקט הייתה לקדם את מחקר הגניזה, להפיץ את פרות המחקר ולהנגיש אותם לציבור הרחב. לשם כך הוקמו צוותי מחקר בתחומים שונים, כאשר בראש כל צוות עומד חוקר בעל שם בתחומו, האחראי על מספר תלמידי מחקר. כל צוות היה אחראי לקטלג, לזהות ולרשום את כל קטעי הגניזה שבתחום אחריותו. כך הוקמו צוותים לפרשנות מקרא, הלכה, מדרש, חכמת הלשון, פילוסופיה ועוד.

פרויקט הדיגיטציה של הגניזה

בשנת 2006 הוקם צוות "גנזים" בראשותו של פרופ' יעקב שויקה. צוות "גנזים" היה אחראי להקים ולתחזק את התשתית הטכנולוגית שתרכז את כל התוצאות של מחקר הגניזה בעבר ובהווה, ותעמיד אותו לרשות החוקרים והציבור הרחב באמצעות אתר אינטרנט ייעודי. אחת המשימות הראשונות שהוטלו על "גנזים" הייתה להכין אינוונטר מלא וראשון מסוגו של כל קטעי הגניזה בעולם, לבצע דיגיטציה של כל קטעי הגניזה, וליצור באמצעות אתר האינטרנט את הספרייה הוירטואלית של הגניזה הקהירית. הכנת האינוונטר הושלמה זה מכבר, וכיום יש בידינו לראשונה רשימה מלאה של כל קטעי הגניזה המצויים כאמור בקרוב לשבעים ספריות ואוספים, מספריית קמברידג' בה מצויים כ-60% מכל הקטעים, ועד לאוספים המכילים קטעים בודדים, ולעיתים אף לא יותר מקטע אחד. יחד עם הכנת האינוונטר התחלנו בצילום דיגיטלי של כל הקטעים. צלמים נשלחו לכל ספרייה שהחזיקה אוסף של קטעי גניזה, קטן כגדול. חוזים משפטיים נחתמו עם כל ספרייה וספרייה כך שתתאפשר הצגת הקטעים לציבור הרחב באמצעות אתר האינטרנט של הפרויקט. גם פרויקט גדול זה עומד לפני סיומו המוצלח. נכון להיום מצויים באתר האינטרנט של הפרויקט למעלה מ-333,000 תמונות של קטעי הגניזה, המקיפים את רוב ככל האוספים הקיימים. הצילום של אוסף הגניזה הגדול שבספריית קמברידג' עדיין לא הסתיים, והוא מתקדם בקצב של כ-10,000 תמונות חדשות כל חודש. גם הצילום של אוסף הספרייה הבריטית נמצא בעיצומו. מלבד סיום שני אוספים אלו אנו חסרים עדיין את האוספים של ספריית אוקספורד, ספריית מנצ'סטר והספרייה בסנט פטרבורג, ואנו מקווים להגיע בעתיד הקרוב לסיכום מוצלח גם עם ספריות אלו.

השוואה לספריות וירטואליות אחרות

מאגר גדול זה של תמונות דיגיטליות מעמיד את האתר של פרויקט פרידברג כספרייה הוירטואלית הגדולה ביותר בעולם של כתבי-יד. להלן השוואה עם פרוייקטים דומים:

- APIS = Advanced Papyrological Information System של אוניברסיטת קולומביה מרכז מידע ותמונות דיגיטליות של פפירוסים מספריות שונות בעולם. נכון לסוף 2009 הוא הכיל 34,000 תמונות בסך הכל (אין מידע מעודכן יותר).
- IDP = The International Dunhuang Project של הספרייה הבריטית בשיתוף כעשרים מוסדות נוספים, מכיל תמונות ומידע על כתבי-יד מהמאה הראשונה לספירה ועד המאה ה-14, שנתגלו ברובם במערה בעיר Dunhuang בקצה מדבר גובי שבמערב סין, כמו גם בערים אחרות על דרך המשי. זהו הפרויקט הקרוב ביותר מבחינת אופיו לפרויקט פרידברג. נכון לנובמבר 2011 הפרויקט מכיל כ-323,000 תמונות.
- Digital Scriptorium – מרכז תמונות ומידע של כתבי-יד מימי הביניים ומהרנסאנס מספריות דיגיטליות בארצות הברית. 24,300 תמונות בסך הכל.
- E-codices: Virtual Manuscript Library of Switzerland – 833 כתבי-יד. המשותף לכל הפרויקטים הנ"ל הוא שבשום מקרה לא נטל על עצמו מוסד אחד לצלם כתבי-יד המפוזרים במקומות השונים, ולכל היותר נבנתה תשתית המאפשרת גישה לאוספים שונים דרך פורטל אחד. גם בזה פרויקט פרידברג מתייחד וקובע מקום לעצמו.

היבט נוסף בו מתייחד פרויקט פרידברג משאר הפרויקטים הקיימים, הוא בתנאי הצילום. לפני כשנתיים התחלנו בפרויקט במחקר על ניצול של טכנולוגיות של ראייה ממוחשבת לעיבוד אוטומאטי של התמונות, כמפורט בהמשך. אחת המסקנות הראשונות שלנו היו כי על מנת להגיע לתוצאות הטובות ביותר, יש לדאוג לתנאים מסוימים שיתקיימו בצילום הדיגיטלי. בין התנאים העיקריים: צילום על רקע כחול על מנת שתהיה הפרדה מקסימלית בין הרקע לכתב-היד, וצילום עם סרגל בכל תמונה, על מנת שיהיה אפשר לשחזר באופן אוטומאטי את הרזולוציה המדויקת של התמונה. תנאים אלו, בפרט הצילום על רקע כחול, הנם ייחודיים לפרויקט פרידברג. אי-הקפדה על תנאים אלו, כפי שמצוי בספריות דיגיטליות רבות, גורמת לכך שניסיון עתידי לעבד את התמונות הדיגיטליות של כתבי היד ייתקל בקשיים מיותרים ויפגע באיכות התוצאות.

*

ניצול טכנולוגיות האינטרנט בפרויקט פרידברג מקדם את מחקר הגניזה בשני היבטים עיקריים: נגישות לקהילת החוקרים והקהל הרחב במידה שלא הייתה קיימת בעבר, ואפשרות לשחזר כתבי-היד ועיון בהם בצורתם היותר מקורית.

נגישות

עד שנפתח אתר פרידברג לחקר הגניזה, חוקרים שרצו לעיין בקטעים היו צריכים לנסוע אל הספריות השונות על מנת לצפות בקטעים במקורם, או להשתמש בצילומי המיקרופילם הנמצאים במכון לתצלומי כתבי-יד של הספרייה הלאומית בירושלים. כמובן שנסיעה לספרייה באירופה או בארצות הברית אינה אפשרית בכל עת לכל אחד, נוסף על כך שרבות מהספריות מגבילות את העיון בכתבי-יד לחוקרים בלבד. תהליך הזמנת כתבי-יד כרוך גם הוא בזמן המתנה של מספר שעות. מכיוון שרבים מקטעי הגניזה אינם מקוטלגים עדיין, והחוקר נדרש במקרים רבים לעבור על אוספים שלמים על מנת לזהות קטעים הדרושים למחקרו, האפשרות להזמנת כתבי-היד בודדים בספרייה עצמה אינו עונה על הצורך בסריקה מהירה של האוסף כולו.

השימוש בצילומי המיקרופילם בספרייה הלאומית אכן מאפשר סריקה מהירה של אוספים שלמים, אולם כאן החוקר מוגבל לאיכות הבינונית של הסרטים המצולמים בשחור ולבן, ורובם כבר שרוטים ופגומים מרוב שימוש. כמובן, גם כאן החוקר מוגבל לעבודה במקום מסוים, בשעות בהן הספרייה פתוחה, ובתנאים שאינם אופטימליים.

אתר פרידברג לחקר הגניזה מאפשר בפעם הראשונה צפייה בתמונות צבעוניות ברזולוציה גבוהה של כל קטעי הגניזה, מכל מקום בעולם ובכל זמן. ואכן, 1,500 החוקרים הרשומים באתר מעידים כי הרגלי העבודה והאפשרויות למחקר השתנו עם פתיחת האתר לבלי היכר. נוסף על כך, ריכוז המידע הקטלוגי באתר מעשרות קטלוגים ופרסומים שונים, יחד עם מנוע חיפוש משוכלל המאפשר חיתוך של פרמטרים שונים מעשרות שדות, כל אלו מאפשרים גישה ישירה לקטעים הרלוונטיים גם ללא כל מידע מוקדם.

יתר על כן, אתר פרידברג פתח את מחקר הגניזה ואוצרותיה עבור כל משכיל ומתעניין, גם אם הוא חסר הכשרה אקדמאית מתאימה. בין המשתמשים הקבועים באתר ישנם רבים מהחברה הדתית והחרדית, בעלי השכלה תורנית אולם ללא גישה לספריות אוניברסיטאיות. עבור קהל זה האתר מהווה את האפשרות היחידה לצפות בקטעי הגניזה, ולהרחיב את השכלתו התורנית במחקר הגניזה של מאה השנים האחרונות. בכך תורם האתר באופן ישיר להנחלת הידע שהיה עד עתה

ספון בהיכלי השן של האקדמיה לציבור הרחב, מעורר את סקרנותם של רבים ופותח בפניהם אופקים חדשים של ידע והשכלה שהיו סגורים בפניהם עד עתה.

צירוף קטעים ושחזור כתב-יד

אופייה של הגניזה גרם לכך שלא נשתמר בה אפילו כתב-יד אחד שלם. בדרך כלל קטע גניזה יהיה דף בודד בלבד, ולמעשה ברוב המקרים מדובר רק בקרעים: חצי דף, רבע דף או קרע קטן של מספר שורות. אמנם הניסיון מלמד שברוב המקרים ניתן למצוא דפים נוספים מאותו כתב יד, ואפילו קרעים מאותו דף, הנמצאים תחת מספרי מדף שונים באותה ספרייה, או אף בספריות שונות. לעיתים רבות שני חצאים של אותו דף נמצאים משני עברי האוקיינוס, לדוגמה חצי דף בקמברידג' שבאנגליה וחציו השני בניו יורק. איתורם של הדפים השונים שהשתייכו במקורם לכתב-יד אחד וצירופם זה לזה, מהווה מרכיב מהותי בעבודתו של חוקר הגניזה. עד עתה הדרך היחידה לעשות דבר זה הייתה על ידי הזמנת צילומים של הקטעים החשודים להיות מתאימים, ובדיקתם זה מול זה. ברור שאי אפשר לעשות השוואה זו במקומם של הקטעים, כאשר כל אחד מהם נמצא פיזית במקום אחר. גם השימוש בצילומי המיקרופילם מוגבל מאוד, כיוון שהוא מחייב שימוש בשתי מקראות בו זמנית, דבר שבדרך כלל אינו זמין. אם שני הקטעים היו במקרה באותו סרט, אי אפשר היה לצפות בשניהם בו זמנית בכל מקרה.

אתר פרידברג שינה מצב זה לחלוטין. מעתה אפשר לצפות בכל זוג קטעים זה ליד זה, ללא תלות במקומם הפיזי. אפשר להשוות על מסך המחשב קטע מירושלים לקטע מפריז או מניו יורק, להגדיל כל קטע, לסובב אותו, לכוונן את הבהירות והניגודיות שלו, ולשפוט האם מדובר בשני קטעים מאותו כתב-יד או לא. מעתה, תחת לחקור קטעים בודדים, החוקר יכול לעיין בכל הדפים שנשתמרו מכתב-יד מסוים בזה אחר זה, ללא תלות במקומם הפיזי כיום.

צירוף קטעים אוטומאטי

אולם מעבר לאפשרות הטכנית לצפות במספר קטעים בו-זמנית, פרויקט פרידברג מעלה תרומה משמעותית ביותר למחקר הגניזה עצמו, באמצעות שימוש בטכנולוגיות מתקדמות של ראייה ממוחשבת. כפי שדיווחנו לאחרונה בעיתונות, פרויקט פרידברג בשיתוף עם חוקרים מהחוג למדעי המחשב באוניברסיטת תל-אביב הצליח להגיע לפריצת דרך בזיהוי אוטומאטי של קטעים השייכים לאותו כתב-יד, באמצעות השוואתם של כתיבת-יד הסופר בקטעים השונים. טכנולוגיה זו, הדומה בעיקרה לזו המשמשת בזיהוי פנים, מאפיינת את כתב-היד בכל קטע באמצעות ווקטור נומרי. השוואה בין הווקטורים הנומרים של הקטעים השונים מאפשרת לזהות קטעים החשודים כשייכים לכתב-יד אחד. חשוב לציין כי מלאכה זו אינה יכולה להיעשות בשלמות על ידי אדם, ולו רק מפאת גודלה העצום של הגניזה, המכילה למעלה מרבע מיליון קטעים. השוואה של כל הזוגות האפשריים דורשת סדר גודל של רבע מיליון בריבוע פעולות השוואה, שהם עשרות מיליארדים של פעולות. זו היא משימה מאתגרת אף עבור המחשבים החזקים העומדים לרשותנו כיום, ואנו ממשיכים במחקר כדי להבטיח שנוכל לבצע אותה בשלמות ובאמינות בזמן קצר.

מדידות אוטומטיות של קטעים

בצד צירוף הקטעים, הטכנולוגיות של ראייה ממוחשבת משמשים אותנו גם למשימה נוספת: מדידה אוטומטית של פרמטרים שונים של קטעי הגניזה, ביניהם: אורך ורחב של הקטע, גודל השטח הכתוב ורוחב השוליים מכל צד, מספר השורות בקטע ועוד. פרמטרים אלה, הנדרשים בכל קטלוג של קטעי גניזה, מצטרפים לבסיס נתונים המקיף פעם ראשונה את כל קטעי הגניזה. בסיס נתונים כזה יכול לשמש את הקודיקולוגיים והפליאוגרפים החוקרים את מלאכת הכנת הספר והכתיבה בימי הביניים, אולם תרומתו העיקרית תהיה בעזרה באיתור דפים המצטרפים לכתב-יד אחד, האמורים להיות כמובן מתאימים זה לזה גם מבחינת נתונים אלו. נוסף על נתונים אלה אנו מודדים נתונים יותר עדינים, שלא נמדדו מעולם בצורה שיטתית, ביניהם: גובה שורה ממוצע, רווח בין השורות, צפיפות כתב וכדומה. שוב, נתונים אלו יכולים לעזור בתהליך קבלת ההחלטות לגבי צירופי קטעים. מטבע הדברים, ישנו אחוז מסוים של שגיאה במדידות אוטומטיות אלו, הנובע מקטעים כהים מאוד או בהירים מאוד, בהם הדיו אינה נראית ברור, או שהפרדת הקטע מהרקע עליו הוא מצולם נכשלה. ייתכן גם שהקטע צולם בזווית לא ישרה, והיישור האוטומטי של הקטע נכשל. אתר פרידברג נותן אם כן למשתמש כלי גרפי המשרטט על גבי הקטע עצמו את הגבולות שנמדדו, וכך הוא מאפשר למשתמש לשפוט מיד האם המדידה תקינה או לא. חשוב לציין ששרטוטים אלו אינם משנים את התמונה עצמה, אלא מופיעים כשכבה על גביה, ואפשר להדליק ולכבות אותם לפי דרישה.