

הגדרת הפרויקט

בניית מערכת להצגה ולמחקר של כתבי יד עתיקים.

התלמידים יקבלו קבצי וורד שבהם מיוצגים כתבי יד עתיקים של ספר הזהר, וההבדלים בין כתבי יד אלו. התלמידים ימירו את קבצי הוורד לתוך מסד נתונים ויבנה מערכת חיפוש שתשרת חוקרים בתחום מדעי הרוח.

רקע כללי

הנגשת טקסטים קנוניים יהודיים הייתה אחת המשימות שמחקר מדעי היהדות התקופה המודרנית הציב לעצמו, זאת במקביל למחקרם. האתגר הגדול בהנגשה זו הייתה אופיים של טקסטים אלו והאופן שבו הם הגיעו לידינו:

- א. לעומת טקסטים מודרניים, טקסטים אלו לא היו 'יציבים' – כלומר הם התפתחו באופן מקביל באופנים שונים. כלומר, לא מדובר על 'טקסט אחד' אלא על מכלול שלם של טקסטים שעומדים במגע זה עם זה, שכיום כולם התקבצו לכדי 'טקסט' אחד.
- ב. הטקסטים לא השתמרו בכתב יד אחד, אלא הועתקו במשך מאות שונים בכתבי יד.

התוצאה של מאפיים אלו היא כי טקסטים קלאסיים הגיעו אלינו בכתבי יד רבים, השונים מאוד זה מזה, הן ברמת המיקרו (ההבדל בין מילים בתוך משפט) והן ברמת המאקרו (המצאי הכללי, המבנה הכללי של החיבור).

ספר הזהר הינו ספר היסוד של תורת הסוד היהודית. על פי המחקר המודרני, הספר נכתב בספרד לאורך ימי הביניים, ונדפס לראשונה באיטליה במאה השש-עשרה. הספר מורכב מיחידות שונות, שהסדר שלהם, באופן כללי, הוא על סדר התורה (פרשות שבוע). עוד לפני הדפוס, הספר מצוי בידינו בעשרות כתבי יד חלקיים, הנבדלים ביניהם הן במצאי, אך גם בנוסח הטקסט שלהם.

במסגרת 'פרויקט הזהר' בראשות פרופ' רונית מרוז, הוכנה מהדורה סינופטית בדפוס (כלומר – כזו המציגה כל נוסח כתב יד זה לצד זה) של חלקים ממסוימים מתוך ספר הזהר. על הפרויקט ועל הפרשות אפשר לקרוא במבוא למהדורה (בקישור למטה). המהדורה המודפסת (שלא יצאה לאור בספר, אלא רק בקבצי PDF) המצורפים. הטקסטים קיימים בידינו גם כקבצי וורד) משקפת את האתגר הגדול בהצגה של טקסט זה באופן מודפס.

הפרויקט הוא הצעד הראשון לקראת מהדורה מדעית של החיבור, כאשר הוא מביא בצורה מפורטת את כל כתבי היד שבידינו על יחידות אלו. מטרת הפרויקט הדיגיטלי, היא לקחת את המידע אשר נאסף בפרויקט, לפענח אותו מתוך הקבצים ולהציג אותו באופן שיאפשר מיצוי של הידע המצוי בו על ידי חוקרי התחום.

מטרת הפרויקט:

- המרה של מהדורת הדפוס למהדורה דיגיטלית.
- יצירת ממשקים המאפשרים הסתכלות על הטקסטים באופנים מורכבים (כפי שמהדורות דיגיטליות מאפשרות), כולל חלוקה מובנית ליחידות טקסט, הצגה של כתבי יד לפי אשכולות, ממשק לשאילתות, ויזואליזציה של תוצאות.

קבצים נלווים: [קישור לקובץ PDF](#). קיימים גם קבצי Word מקבילים שינתנו לסטודנטים שיעבדו על הפרויקט.

תהליך העבודה:

- (1) לימוד של מבנה המהדורה המודפסת (דורש הבנה במהדורות אקדמאיות לספרות קלאסית).
 - הבנה של מהי מהדורה מדעית. מה המשמעות של יחסים בין טקסטים. הנחיה צמודה של ד"ר אליעזר באומגרטרן (מנהל מעבדת אליהו).
- (2) פענוח של מבנה קבצי wordn דרך המרתם לקבצי XML.
 - שימוש בתיקיית פייטון. דורש פענוח של דרך ההצגה של הסימונים השונים של קבצי הוורד, והאופן שהם מיוצגים בתוך הXML.
- (3) יצירת סכמה למבנה חדש של המהדורה הדיגיטלית.
 - תהליך אפיון מול חוקרים. UX/UI.
 - יצירת סכמת מידע, בחירת מסד נתונים כו'.
- (4) המרה של קבצי XML למבנה החדש.
 - לאור סכמת הידע שנבחרה, צריך להמיר את מבנה הנתונים הישן למבנה נתונים חדש (כתיבת קוד, או דרך אחרת).
- (5) יצירת מסד נתונים והעברה של המידע למסד הנתונים.
- (6) בנייה של ממשקים לקריאה של התוצאות, לשאילתות.

איש הקשר: אליעזר באומגרטרן beliezer@staff.haifa.ac.il