

Project in Data & Knowledge (236323) – Winter 2016-2017

A Semantic Knowledge Base for the Jewish Bookshelf (JBS)



amee rivers @ flickr

המעבדה למידע וידע (TDK) מזמינה אתכם לקחת חלק בפרויקט חדשני: יצירת מאגר ידע מבוסס Semantic Web עבור **ארון הספרים היהודי**, אשר יאפשר שאלות מדויקות ועשירות במיוחד. כיום, קיימים אמנם מספר מאגרי מידע ממוחשבים לספרות היהודית דוגמת שו"ת בר-אילן (אשר זכה בשנת 2007 בפרס ישראל), אך מאחר וחסר שם ייצוג סמנטי של המידע, יכולות החיפוש שלהם מוגבלות מאוד ומבוססות בעיקר על

מילות-מפתח. בנוסף, האינטגרציה עם מערכות תוכנה אחרות מוגבלת גם היא.

הסטודנטים אשר ישתתפו בפרויקט יחשפו לטכנולוגיות המרתקות של ה-Semantic Web אשר הוגדרו על ידי ממצאי ה-Web, טים ברנרס-לי, כשלב הבא באבולוציה של ה-Web. הטכנולוגיות הללו מאפשרות להגדיר קשרים סמנטיים בין ישויות מידע שונות ליצירת גרף ידע עתיר-מימדים. אחד מהפרויקטים המפורסמים בתחום זה הוא **DBpedia** (dbpedia.org) שם מפתחים מאגר ידע סמנטי על סמך מידע הנשלף מהאנציקלופדיה המקוונת Wikipedia באופן אוטומטי.

ארון הספרים היהודי מכיל כמות גדולה מאוד של טקסטים בעברית אשר ביניהם קשרים רבים ומגוונים. מטרת הפרויקט היא לחשוף את הקשרים הנ"ל (**באופן אוטומטי**) ולייצגם בצורה נכונה בטכנולוגיות של ה-Semantic Web. בשלב הראשון אנו נתמקד בעיקר בקשרים מסוג relates-to אשר יאפשרו למשל את השאלות הבאות:

- (1) מהם המקומות בתלמוד הבבלי ובתלמוד הירושלמי אשר בהם יש התייחסות ל"חזון העצמות היבשות" של הנביא יחזקאל? באלו מקומות בתלמוד יש התייחסות לנושאים הקשורים למדעים מדויקים?
- (2) מהם הפסוקים בספר שמות אשר בפירושם מתייחס הרמב"ן אל הרמב"ם? באלו הפסוקים הללו הרמב"ן חולק על הרמב"ם?

הפרויקט יינתן במסגרת הקורס 236323 (3 נקודות) ובו יפתחו הסטודנטים בשפת התכנות Python, וישתמשו בטכנולוגיות של ה-Semantic Web דוגמת RDF, RDFS, OWL, ושפת השאלות SPARQL. בנוסף, ייושמו בפרויקט טכניקות של חילוץ מידע (Information Extraction) למשל עבור זיהוי שמות אישים בטקסט (Named Entity Recognition). העבודה תתבצע בצוותים של 2-3 סטודנטים.

מספר המקומות מוגבל, ההרשמה בתיאום עם מנחה הפרויקט ד"ר אורן משלי
omishali@cs.technion.ac.il

דרישות קדם:

- לפחות שנה שלישית, קורס במערכות מסדי נתונים – יתרון. הפרויקט פתוח גם לסטודנטים מתואר שני.

מקורות:

- <https://www.w3.org/standards/semanticweb/>: Semantic Web
- מאמר על ארכיטקטורת DBpedia: <http://www.semantic-web-journal.net/system/files/swj499.pdf>
- עקרונות בחילוץ מידע: <http://benny.net.technion.ac.il/files/2015/02/pods14-tutorial-kimelfeld.pdf>