

ניהול נתוני עתק ברשת בעזרת ההמונים

Web, Crowd and Big Data Management

89-452-01

Course type: seminar

סוג הקורס: סמינריון

Extent: two hours per week

היקף שעות: שעתיים בשבוע

Website: [MOODLE](#)

אתר הקורס באינטרנט: [MOODLE](#)

א. **מטרות הסמינריון:** נדון בנושאים עכשוויים בחזית המחקר בניהול נתונים (data management), ובפרט באתגרים הנוגעים לניהול וניתוח נתוני עתק (big data), לנתונים שמגיעים ממקורות שונים באינטרנט (web data) ולנתונים אותם נרצה לאסוף מהמוני משתמשי האינטרנט או לעבד בעזרת אותם משתמשים.

Outcomes: we will introduce current topics in data management research, focusing on challenges related to big data frameworks, Web data and the harvesting or processing of data with the crowd of Web users, namely crowdsourcing techniques.

ב. **תוכן הסמינריון:** הסמינריון יתבסס על הרצאות של הסטודנטים, בנושאים מתקדמים. הסטודנטים ידרשו לבחור מאמר ולהציג אותו, ביחידים או בזוגות (כתלות במספר הנרשמים). בכל שבוע יוצג מאמר אחד במהלך תשעים הדקות של השיעור.

Contents: the seminar is based on student presentations. Students are required to select one research paper and present it to their peers in class, in singles or couples (depending on the number of registered students). A single paper will be presented each week, over the ninety minutes of class.

The papers will involve the following topics: המאמרים יהיו בנושאים הבאים:

Big data management: MapReduce, architectures, data exploration

Web data management: Semantic Web, Web-based knowledge bases, SPARQL, information extraction from semi-structured data

Crowdsourcing: data harvesting and analysis with the help of the crowd

ג. חובות הקורס:

דרישות קדם: מעבר קורסי חובה שנה א' וב', מבני נתונים, רצוי מערכות מסדי נתונים (או מסדי נתונים)

חובות / דרישות / מטלות: הצגת מאמר בודד או שני מאמרים בזוג, כתלות במספר הנרשמים. חובת נוכחות בכל ההרצאות.

מרכיבי הציון הסופי (ציון מספרי / ציון עובר): 95% הרצאה, 5% על השתתפות פעילה בדיון בהרצאות של אחרים.

Requirements:

Prerequisites: first and second year courses, optionally database systems (89-851) or databases (89-281)

Assignments and other duties: presenting in class, in singles or couples (depending on the number of registered students). Attendance is mandatory.

Grade: 95% presentation, 95% active participation in discussion during peers' presentations.

ג. **ביבליוגרפיה:** רשימת מאמרים מומלצים תתפרסם באתר הקורס בתחילת הסמסטר.

Bibliography: the list of recommended papers will be announced at the beginning of the semester, at the course Website.

Recommended reading:

Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." *Communications of the ACM* 51.1 (2008): 107-113.

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific american* 284.5 (2001): 28-37.

Doan, Anhai, Raghuram Ramakrishnan, and Alon Y. Halevy. "Crowdsourcing systems on the world-wide web." *Communications of the ACM* 54.4 (2011): 86-96.

Parameswaran, Aditya, and Neoklis Polyzotis. "Answering queries using humans, algorithms and databases." (2011).