

האוניברסיטה העברית בירושלים

סילבוס

כריית מידע מנתונים גדולים - 52002

תאריך עדכון אחרון 27-08-2024

נקודות זכות באוניברסיטה העברית: 3

תואר: בוגר

היחידה האקדמית שאחראית על הקורס: סטטיסטיקה

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר א'

שפת ההוראה: עברית

קמפוס: הר הצופים

מורה אחראי על הקורס (רכז): אור צוק

דוא"ל של המורה האחראי על הקורס: or.zuk@mail.huji.ac.il

שעות קבלה של רכז הקורס: שני 10:30-11:30

מורי הקורס:

ד"ר אור צוק

תאור כללי של הקורס:
נלמד שיטות לטיפול ולניתוח מסדי נתונים גדולים.

מטרות הקורס:
רכישת כלים תכנותיים וסטטיסטיים הדרושים לביצוע סטטיסטיקות על נתוני ענק.

תוצרי למידה
בסיומו של קורס זה, סטודנטים יהיו מסוגלים:
לנתח מסדי נתונים המכילים מליוני רשומות ואלפי משתנים. להשתמש בצורה יעילה בתוכנות המסוגלות לבצע חישוב מקבילי/בענן. להוציא ולנתח נתונים המרוכזים ברשתות.

דרישות נוכחות (%) :
0

שיטת ההוראה בקורס: הרצאה, הדגמה תוך שימוש במחשב.

רשימת נושאים / תכנית הלימודים בקורס:
עבודה מרחוק בסביבת אוח ובמחשוב ענן, SQL, אחזור מידע מהרשת.
מציאת דמיון: מציאת שכנים, פונקציות האש
חישוב מבוזר בסביבת ענן
ניתוח רשתות: זיהוי קהילה, דגימה בגרפים גדולים.
זרם נתונים: אלגוריתמי און-ליין, קבלת החלטות און-ליין (מבחינת B/A)

חומר חובה לקריאה:
אין

חומר לקריאה נוספת:
Leskovec, Rajaraman&Ullman (2014). Mining of massive datasets, Cambridge University Press

Tan, Steinbach, Karpatne and Kumar (2005). Introduction to Data Mining. Pearson Addison Wesley

-Liu (2011). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data Centric Systems and Applications). Springer

White (2015). *Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale*.
O'Reilly Media

מרכיבי הציון הסופי :

הגשת עבודה מסכמת / פרויקט גמר / מטלת סיכום / מבחן בית / רפרט 75 %
מטלות הגשה במהלך הסמסטר: תרגילים / עבודות / מבדקים / דוחות / פורום / סימולציה ואחרות
25 %

מידע נוסף / הערות:

ינתנו במהלך הסמסטר עבודות/ות ביניים וכן עבודות list-check קצרות שיהוו ביחד 25% מציון הקורס.
לאחר סוף הסמסטר תהיה עבודת בית שתהווה 75% מציון הקורס.
כמו כן יינתנו מספר תרגילי רשות לתרגול חומר הקורס.