
האוניברסיטה העברית בירושלים

סילבוס

מחט בערימת דאטה: מבוא למדע המידע - 67978

תאריך עדכון אחרון 27-09-2024

נקודות זכות באוניברסיטה העברית: 3

תואר: בוגר

היחידה האקדמית שאחראית על הקורס: מדעי המחשב

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר ב'

שפת ההוראה: אנגלית

קמפוס: קרית א"י ספרא

מורה אחראי על הקורס (רכז): פרופ' דפנה שחף וד"ר תום הופ

דוא"ל של המורה האחראי על הקורס: dshahaf@cs.huji.ac.il, tomhope@cs.huji.ac.il

שעות קבלה של רכז הקורס: TBA

מורי הקורס:

תאור כללי של הקורס:

מדע המידע (Science Data)

הוא תחום העוסק בזיהוי תבניות ודפוסים במידע גולמי.

הצמיחה בנפח המידע הדיגיטלי הגבירה את הצורך בשיטות אוטומטיות לניתוח מידע; שיטות אופייניות משלבות סטטיסטיקה, למידת מכונה ומסדי נתונים. למדע המידע מגוון רחב של יישומים, ממדע וטכנולוגיה דרך תעשייה, חברה וכלכלה.

מטרות הקורס:

תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

להבין איזה סוג כלים (ונתונים) הם צריכים בשביל לגשת לבעיה.

דרישות נוכחות (%) :

0

שיטת ההוראה בקורס:

רשימת נושאים / תכנית הלימודים בקורס:

רשימת נושאים טנטטיבית:

* רקע שימושי: היסק סטטיסטי

* דמיון ומרחק

- מרחקים בטקסט, עיבוד טקסט, *learning contrastive, embedding text*

* מודלי שפה

- BERT, T5, GPT

* חילוך מידע מובנה מטקסט

* אשכולות (קלאסטרים)

* גרפים

- רשתות חברתיות, זיהוי קהילות, משולשים, רשתות עולם קטן, *link, embedding graph*

prediction

* מערכות המלצה

* כריית מידע אינטראקטיבית, ויזואליזציה

* תכנון ניסויים

- ניסויים אקראיים לעומת תצפיתיים, סיבתיות

* MapReduce, Hadoop

* הפחתת מימדים

* זרמי נתונים

חומר חובה לקריאה:
TBA

חומר לקריאה נוספת:

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
15 %
מבחני אמצע 10 %

מידע נוסף / הערות:

דרישות: קורס לתארים מתקדמים ושנה שלישית של תואר ראשון. ידע של שפת תכנות ו/או שפת
scripting.
ידע בסיסי באלגוריתמים והסתברות. מאוד מומלץ לקחת ML קודם.