

סילבוס לקורס: ניתוח טקסטים בכלים דיגיטליים וחישוביים

מספר: 1886.4001

היקף: 2 ש"ס

קורס סמסטריאלי

מיועד לסמסטר א' תשפ"ו לתלמידי מחקר בפקולטות למדעי הרוח, החברה, משפטים.

מרצה: ד"ר ורד זילבר-ורוד (מרכז TAD) בשיתוף עם מרצים אורחים.

עוזרת הוראה: גב' סתיו קליין

הרכב הציון – הגשת תרגילים (20%) ומטלת סיום קורס (80%).

יום ושעה: ימי שלישי בין השעות 14:15 - 15:45

תיאור:

הקורס מתמקד ביסודות תחום ניתוח חישובי של טקסטים ובהכשרת התלמידים לעבודה עצמאית בכלים דיגיטליים לניתוח טקסטים.

המחקר המסורתי בדיסציפלינות השונות במדעי הרוח מעוגן בצורות חשיבה מסוימות. לטקסט תפקיד מרכזי בתרבות האנושית, מן הטקסטים הקדומים ועד ספרות ועיתונות בת זמננו. בעידן שבו אוספים טקסטואליים שהיו עד כה מונחים על מדפי הספרים הופכים לדיגיטליים ונגישים לעיבוד במחשב האישי, מתרחש מעבר מקריאה צמודה של טקסטים נדירים ל"קריאה רחוקה" של מאגרים רחבי-היקף. בתהליך זה, כל גוף ידע טקסטואלי, קורפוס, הופך באחת לנתונים הניתנים לפירוק ועיבוד וכתוצאה מכך להתבוננות מזווית שונה המאפשרת פרשנות מחודשת. הקורס מתאר את ההזדמנות הייחודית הטמונה בנקודת הזמן הנוכחית לשינוי פני מדעי הרוח המסורתיים שבה כלי GenAI מהווים את הדור הבא של כלים לניתוח. הקורס יכשיר את התלמידים לעריכת מחקרים חישוביים בסיסיים בכוחות עצמם באמצעות תרגול ניתוח טקסטים בכלים לניתוח קורפוס טקסטואלי וכן בכלי AI הזמינים לכול. יחד עם זאת, תוצרי הניתוח הדיגיטלי ייבחנו בצורה ביקורתית, תוך בחינת האתגרים המלווים את תהליך עיבוד הנתונים הטקסטואליים וחשיבות השמירה על עקרונות המחקר המדעי.

- שיעור 1: מהי מילה? ומהי "תמנית"? מהו קורפוס טקסטואלי? תרגול: יצירת רשימות מילים.
- שיעור 2: מבוא לעיבוד שפה טבעית: מאלליזה לג'מיני; ממבחן טיורינג לבינה מלאכותית.
- שיעור 3: התמרת הטקסט: למטיציה, ניתוח מורפולוגי, תיוג חלקי הדיבר, זיהוי ישויות. כיצד הידע הבלשני מסייע לנתח מגמות ודפוסים לשוניים של טקסטים שונים. תרגול: התמרות ותיוג בכלים דיגיטליים.
- שיעור 4: ניתוח קורפוס א' תרגול: הפקת תוצרי ניתוח ב-Antconc
- שיעור 5: ניתוח קורפוס ב' תרגול: הפקת תוצרי ניתוח ב-Voyant Tools
- שיעור 6: שימוש במאפיינים לשוניים כדי להשוות טקסטים: מודלי שפה ושיכון מילים. תרגול: השוואה של מסמכים.
- שיעור 7: זיהוי סגנונות כתיבה (סטילומטריה) תרגול: השוואה של מסמכים.
- שיעור 8: מידול נושאים (Topic modeling) – האם וכיצד ניתן לזהות נושאים שנמצאו במסמכים מבלי לקרוא אותם?
- שיעור 9: רשתות סמנטיות: קשרים בין טקסטים ובין מסמכים. תרגול: יצירת רשת סמנטית.
- שיעור 10: LLM זה לא מפחיד אותי – איך ניתן לרתום כלים מבוססי מודלי שפה גדולים למחקר? תרגול: חיפוש וניתוח בכלי בינה מלאכותית (NotebookLM, Elicit, Citations, etc.).
- שיעור 11: מאפיינים ייחודיים בניתוח טקסט שמקורו בדיבור - עקרונות ותהליכי עבודה. תרגול: מאודיו לטקסט – תמלול אוטומטי ושיוך לדוברים.
- שיעור 12: דוגמאות למחקרים במדעי הרוח שעשו שימוש בשיטות שנלמדו עד כה. תרגול: שימוש ב-AI לויזואליזציה וציטוט. יעשה שימוש אך ורק במערכות המתבססות על עבודה בלעדית עם המסמכים המקוריים שהוזנו אליהן, תוך מתן הפניות מדויקות למקורות.
- שיעור 13: סיכום והצגת הצעות לפרויקטים.

ביבליוגרפיה ראשונית:

1. **מילים שקולות – צעדים ראשונים במחקר הספרות החישובי** / איתי מרינברג-מיליקובסקי. הוצאת למדא האוניברסיטה הפתוחה (2022).
2. **מחקר חישובי במדעי הרוח – אסופת מאמרים** / אופיר מינץ-מנור, איתי מרינברג-מיליקובסקי. הוצאת למדא האוניברסיטה הפתוחה (2022).

3. William J. Turkel and Alan MacEachern, *The Programming Historian* 1st edition (Network in Canadian History & Environment: 2007–2008).
<https://programminghistorian.org/en/about>