

ניתוח טקסטים בכלים דיגיטליים וחישוביים

קורס חדש (1886.4001)

קורס סמסטר־אלי לתלמידי מחקר בפקולטות למדעי הרוח, החברה ומשפטים

מרצה: ד"ר ווד זילבר-ורוד, המרכז לבניה מלאכותית ומדעי הנתונים (מרכז TAD)
בשיתוף עם מרצים אורחים

אנו שמחים להשיק קורס שיתמקד ביסודות תחום ניתוח חישובי של טקסטים ובהכשרת התלמידים לעבודה עצמאית בכלים דיגיטליים ובשיטות חישוביות לניתוח טקסטים.

המחקר המסורתי מבוסס טקסטים בדיסציפלינות השונות מעוגן בצורות חשיבה מסוימות. לטקסט תפקיד מרכזי בתרבות האנושית, מן הטקסטים הקדומים, עיתונות בת זמננו ועד לרשתות חברתיות. 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, המכונה "קורפוס", הופך באחת לנתונים הניתנים לפירוק ועיבוד וכתוצאה מכך להתבוננות מזווית שונה המאפשרת פרשנות מחודשת. הקורס ידגיש גם את ההזדמנות הייחודית הטמונה בנקודת הזמן הנוכחית שבה כלי GenAI מהווים את הדור הבא של כלים לניתוח.

הקורס יכשיר את התלמידים לעריכת מחקרים מבוססי טקסטים בשיטות חישוביות בכוחות עצמם באמצעות תרגול בכלים קיימים, הבנת אלגוריתמים לניתוח קורפוס טקסטואלי וכן בכלי AI הזמינים לכול. יחד עם זאת, תוצרי הניתוח הדיגיטלי ייבחנו בצורה ביקורתית, תוך למידת האתגרים המלווים את תהליך עיבוד הנתונים הטקסטואליים וחשיבות השמירה על עקרונות המחקר המדעי.



סילבוס מורחב

- שיעור 1: מהי מילה? ומהי "תמנית"? מהו קורפוס טקסטואלי?
תרגול: יצירת רשימות מילים.
- שיעור 2: מבוא לעיבוד שפה טבעית: מאליוזה 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: שיכון מילים (word embeddings)
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?
רשתות סמנטיות: קשרים בין טקסטים ובין מסמכים.
- שיעור 9: 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.

היקף: 2 ש"ס

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

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

לפרטים ושאלות: חנה שמילוביץ ממרכז TAD <hanashmil@tauex.tau.ac.il>

רשימת מקורות ראשונית:

1. **מילים שקולות – צעדים ראשוניים במחקר הספרות החישובי** / איתי מרינברג-מילקובסקי. הוצאת למדא האוניברסיטה הפתוחה (2022).
2. **מחקר חישובי במדעי הרוח – אסופת מאמרים** / אופיר מינץ-מנור, איתי מרינברג-מילקובסקי. הוצאת למדא האוניברסיטה הפתוחה (2022).
3. Zef Segal; Nurit Melnik (2025). *Computational Literacy for the Humanities – Mathematics and Programming in Context*. London: Routledge.
<https://doi.org/10.4324/9781003502814>
4. William J. Turkel and Alan MacEachern, *The Programming Historian* 1st edition (Network in Canadian History & Environment: 2007–2008).
<https://programminghistorian.org/en/about>