



אוניברסיטת בן-גוריון בנגב
הפקולטה למדעי הטבע - המחלקה למדעי המחשב
סמסטר א' תשפ"ב

סילבוס קורס

| | |
|------------------|--|
| שם קורס: | עיבוד שפה טבעית |
| שם קורס באנגלית: | Natural Language Processing |
| מספר קורס: | 202-2-5211 |
| סוג קורס: | בחירה |
| נק"ז: | 4.0 |
| מרצה הקורס: | דר' יובל פינטר |
| דרישות קדם: | תכנון אלגוריתמים (20212041), עקרונות שפות תכנות (20212051) |

סילבוס בעברית: (ייתכנו שינויים)

עיבוד שפה טבעית הוא תחום הנסוב סביב השאלות האם וכיצד ניתן להבין ולייצר באופן אוטומטי טקסט שנכתב בשפה שנועדה לבני אדם. בעיות כמו תרגום אוטומטי, זיהוי ישויות, הסקת מסקנות ומענה על צרכי משתמש בציט שהגדירו את התחום במשך שנים ארוכות חוו לאחרונה פריצות דרך משמעותיות הודות לכח החישוב הגובר וכמויות המידע הזמינות לנו, והתחום הפך למוקד מרכזי בגלל הבינה המלאכותית (AI).

כל יחידה בחלקו המרכזי קורס תתחיל בשאלה הנוגעת לאופן השימוש שלנו בשפה ותתמקד בטכניקה חישובית בולטת המותאמת לאותה בעיה. סיכום הקורס יגע בהשלכות הרחבות של פרישת עוד ועוד מערכות עיבוד שפה אוטומטיות לשימוש בעולם הדיגיטלי וה"אמיתי".

מבוא לשפה, למידת מכונה ועיבוד שפה

עיבוד טקסט והגדרת הבעיה

סמנטיקה ברמת המסמך וסיווג

תכונות מילים ותיוג רצפים

תחביר ומידול מבנים

סמנטיקה ברמת המילה ומודלים התפלגותיים

תרגום מכונה וחילול טקסט



אוניברסיטת בן-גוריון בנגב
הפקולטה למדעי הטבע - המחלקה למדעי המחשב
סמסטר א' תשפ"ב

רשתות נוירונים בעיבוד שפה ומודלי שפה גדולים
אתיקה, הסברנות, הכלה וסביבה בעיבוד שפה
יישומי עיבוד שפה בבנייה מלאכותית, מדעי הרוח והחברה, פיננסים ועוד.

[סילבוס באנגלית:](#)

Introductions to Language, Machine Learning, and NLP
Text Processing and Task Definition
Document Semantics and Classification
Word-level Characteristics and Sequence Labeling
Syntax and Structural Modeling
Lexical Semantics and Distributional Modeling
Machine Translation and Text Generation
Neural Nets in NLP and Large Language Models
Ethical NLP, Explainability, Inclusive NLP, Green NLP
Applications of NLP in AI, Humanities, Social Science, Finance, etc.

[דרישות ומרכיבי ציון הקורס:](#)

הקורס מדגיש את הבעיות בתחום עיבוד השפה מזווית בלשנית ומתוך עניין בהשלכות שיש לכלים השונים על התעשייה ועל החברה ולכן פונה בעיקר לסטודנטים בעלי עניין בין-תחומי.

10% - בחנים שבועיים קצרים.

40% - ארבעה תרגילי בית שחלקם יכולול תכנות בשפת פייתון.

50% - מבחן מסכם.

[ספרות הקורס:](#)

- **Daniel Jurafsky and James Martin. Speech and language processing.**
3rd Edition (2020)

בנוסף יילמד חומר ממקורות ברשת וממחקר עדכני.



אוניברסיטת בן-גוריון בנגב
הפקולטה למדעי הטבע - המחלקה למדעי המחשב
סמסטר א' תשפ"ב

ספרות מומלצת:

- **Jacob Eisenstein. Natural Language Processing (2019). [Notes \(2018\)](#).**
- **Emily M. Bender. Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax (2013)**