



מכון טכנולוגי חולון
Holon Institute of Technology

65339 - NLP-עיבוד שפה טבעית (דיבור וטקסט) בעזרת מחשב - Natural Language Processing using Computers

סמסטר ב', שנה"ל תשפ"א

שם המרצה: ד"ר נאוה שקד

מתרגלים: יורי יורצ'נקו

אופן הוראה: הרצאה + תרגול

שעות שבועיות: 4

נקודות זכות: 3

דרישות קדם: לפחות 3 קורסים מתוכם

61307 תכנות מונחה עצמים- Object Oriented Programming

20021 הסתברות

60073 אלגוריתמים 1

61303 מערכות בסיסי נתונים - Data Base Systems

מטרות הקורס:

NLP הוא מרכיב נדרש היום בכל יישומי הבינה המלאכותית העוסק בממשק אדם מכונה מבוסס שפה. עיבוד שפות טבעיות הוא תחום שמטרתו לאפשר למחשב לנתח שפה טבעית כתובה או מדוברת (text), (or speech) בצורה הדומה למוח האנושי, ומאן ההשקה לתחום הבינה המלאכותית (AI). עוזרים אישיים כגון Siri, Alexa, Cortana כמו גם יישומים בוטים- Bots למיניהם מחייבים מודול כזה על מנת לבצע דיאלוג ומשימות.

מבנה הקורס ותוצריו:

הקורס נועד להציג בפני הסטודנטים את המורכבות במידול שפות אנוש, וכן להקנות ידע מעשי שסייע להם לפתור משימות בתחום עיבוד שפות טבעיות, החל מכלי מיון וכלה בתרגום אוטומטי. בקורס יחשפו הסטודנטים לתיאוריות המדעיות והאסכולות השונות בתחום בתקופות השונות. במעבדה יוצגו שיטות וכלים עדכניים המשמשים לעיבוד שפה ונתרגל עיבוד והקמת בסיסי נתונים, מודלי שפה ופיתרון בעיות עמימות. לכל התיאוריות, האלגוריתמים והכלים נביא דוגמאות אמיתיות של מוצרים וטכנולוגיות שפותחו בתחום ונקשר לפרויקטים אמיתיים בתעשייה.

הקורס בנוי כסמינר, שבו יחקרו הסטודנטים תחומים שונים של מידול שפה ויכולו להתרכז בנושא אחד, אתגר מעניין אותו יחקרו. המחקר של הסטודנט בקורס יעסוק בהיבטים אלגוריתמיים, בנושאים לשוניים ובהדגמת כלים המבצעים ניתוחים על נתוני שפה טבעית. הסטודנט יבנה קשר ישיר בין התיאוריות ליישומים ויקשר מחקר ליישומי תעשייה תוך לימוד מקרים ודוגמאות לפרויקטים בחזית תעשיית ההיי-טק והמחקר.

לאורך הקורס יבנה פרויקט לבחירת הסטודנט, הפרויקט יהיה בעל אופי מחקרי או יישום NLP, המבטא את האתגרים שנלמדו בקורס. התלמידים יעבדו בזוגות ויתבקשו להציג פרזנטציה כיתתית של הפרויקט.

גולומב 52, ת.ד. 305, חולון 58102

טלפון 03-5026528, פקס 03-5026733

52 Golomb St., Holon 58102 Israel

www.hit.ac.il Tel. 972-3-502-6528, Fax. 972-3-502-6733

הפקולטה למדעים

המחלקה למדעי המחשב

Faculty of Sciences

Department of Computer Science



מכון טכנולוגי חולון
Holon Institute of Technology

הנושאים שיילמדו (תוכנית משוערכת לפי שבועות):

שבוע 1:	עיבוד שפות טבעיות <ul style="list-style-type: none"> הגדרה ומושגי יסוד שימושים עיקריים ומטרות שורשים וסקירה היסטורית: מבחן טיורינג, אלייזה
שבוע 2:	עיבוד שפה <ul style="list-style-type: none"> סקירה כללית על עיבוד של רובדי השפה השונים "עמימות": הבעיה, הפתרון, תרגול עיבוד שפה: המוח לעומת המכונה
שבוע 3:	נושאים לשוניים רלוונטיים <ul style="list-style-type: none"> פונטיקה אקוסטית פונולוגיה ופרוזודיה מורפולוגיה לקסיקון
שבוע 4:	נושאים לשוניים רלוונטיים (המשך) <ul style="list-style-type: none"> תחביר וסמנטיקה פרגמטיקה ניתוח שיח
שבוע 5:	מודלי ידע <ul style="list-style-type: none"> אסכולות בעיבוד שפות טבעיות גישות סימבוליות שיטות סטטיסטיות
שבוע 6:	מודלי ידע (המשך) <ul style="list-style-type: none"> גישת "הקשרים" (connectionist) הגישה ההיברידית
שבוע 7:	כלים לעיבוד שפה טבעית <ul style="list-style-type: none"> מבוא לכלים עיקריים על פי קטגוריות לשוניות מושגים מתקדמים מתודולוגיה, כלים ואתגרים
שבוע 8:	מצגות אמצע
שבוע 9:	כלים ואלגוריתמים <ul style="list-style-type: none"> פלטפורמות ליישום מערכות NLP
שבוע 10:	כלים ואלגוריתמים (המשך) <ul style="list-style-type: none"> פלטפורמות ליישום מערכות NLP
שבוע 11:	אינטגרציה של אלגוריתמים וכלים <ul style="list-style-type: none"> אתגרים בעיבוד שפות טבעיות שימוש בכלים לפתרון הבעיות הדגמות לשימוש בכלים בפרויקטים אמיתיים
שבוע 12:	VUI – ממשק למשתמש <ul style="list-style-type: none"> הגדרות ומושגי יסוד תכנון וביצוע שימושיות
שבוע 13:	הגשת פרויקט



מכון טכנולוגי חולון
Holon Institute of Technology

מרכיב הציון:

פרויקט	מצגת	עבודות כיתה וש.ב.	משימה
50%	30%	20%	משקל

Bibliography:

- List of current Articles – will be published.
- Daniel Jurafsky and James H. Martin, *Speech and Language Processing: an introduction to natural language processing, computational linguistics, and speech recognition*, Prentice Hall 2009.
- Oviatt, S., Cohen, P. "The Paradigm Shift to Multimodality in Contemporary Computer Interfaces" Morgan & Claypool Publishers 2015
- Manning, Christopher D., and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT 1999
- Ido Dagan, Dan Roth, Mark Sammons and Fabio Massimo Zanzotto. *Recognizing Textual Entailment: Models and Applications*. Morgan & Claypool Publishers. 220 pages. 2013
- Daniël de Kok, Harm Brouwer *Natural Language Processing for the Working Programmer*, 2011 <http://nlpwp.org/book/> book download
- John E. Kelly III, Steve Hamm. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia Business School Publishing. 2013.

דרישות הקורס:

נוכחות: 80% חובה
נוכחות חובה במצגות כיתה ותרגילים

שעות קבלה: בתאום מראש!

אתר הקורס: אתר ה- moodle