



סילבוס

כריית טקסט וישומים עסקיים - 55703

תאריך עדכון אחרון 20-09-2023

נקודות זכות באוניברסיטה העברית: 3

היחידה האקדמית שאחראית על הקורס: מנהל עסקים

השנה הראשונה בתואר בה ניתן ללמוד את הקורס: 0

סמסטר: סמסטר א'

שפת ההוראה: אנגלית

קמפוס: הר הצופים

מורה אחראי על הקורס (רכז): פרופ רונן פלדמן

דוא"ל של המורה האחראי על הקורס: ronen.feldman@huji.ac.il

שעות קבלה של רכז הקורס: יום ב 2-3

מורי הקורס:

פרופ רונן פלדמן

תאור כללי של הקורס:

הקורס נותן סקירה על הטכניקות והישומים המכזיים של תחום כריית הטקסט. נושאים מרכזיים הם

קטלוג מידע, חלוץ מידע, בניית רובוטים לאסוף מידע, וניתוח סנטימנט. יש התמקדות משמעותית סביב שימוש במודלי שפה גדולים וכלי בינה מלאכותית יוצרת

מטרות הקורס:

מטרת הקורס היא להקנות ידע בשיטות לכריית טקסט ולהתנסות בבניית מערכות לנתוח אוספים גדולים של מסמכים. הקורס מתמקד בטכניקות שונות, אלגוריתמים לממוש טכניקות אלו וישומים של הטכניקות בעולם העסקי

תוצרי למידה

בסיומו של קורס זה, סטודנטים יהיו מסוגלים:

לעצב ולבנות מערכת בסיסית לכריית טקסט על מנת לנתח אוסף גדול של מסמכים. סטודנטים יוכלו ליישם מגוון שיטות לקטלוג מידע, חלוץ מידע, בניית רובוטים לאסוף מידע, וניתוח סנטימנט. יכולת שימוש במודלי שפה גדולים בתוך מערכות תוכנה

דרישות נוכחות (%) :

70%

שיטת ההוראה בקורס: הרצאות פרונטליות של החומר הרצאה של הסטודנטים על שמוש בטכניקות של כריית טקסט לפתרון בעיה עסקית בצוע פרויקט מחקר ישומי בקבוצות של 2-3 סטודנטים

רשימת נושאים / תכנית הלימודים בקורס:

Introduction to Text Mining

a. Architecture of Text Mining Systems

2. Term Extraction

3. Text Categorization

a. RIPPER

b. Naive Bayes

c. Sleeping Experts for Phrases

d. SVM

e. KNN

f. Online Methods

g. Committees

h. Begging and Boosting

4. Information Extraction

a. General Architecture

b. HMM

c. Knowledge Based Systems

d. Boot Strapping

e. Unsupervised relation Extraction

5. Analytics

-
- a. Maximal Association Rules
 - b. Trend Analysis
 - c. Distribution Analysis
 - d. Comparing Profiles
 - 6. Link Analysis
 - a. Pajek
 - 7. Sentiment Analysis
 - a. Document Level Sentiment Analysis
 - b. Sentence level Sentiment Analysis
 - c. Aspect based Sentiment Analysis
 - d. Comparative Sentiment Analysis
 - e. Sentiment Analysis Applications
 - 8. Visualizations
 - a. Circle Graphs
 - b. Spring Graphs
 - c. Trend Graphs
 - 9. Applications
 - a. Content Management
 - i. Classification of documents
 - ii. Automatic organization of internet content
 - iii. Clustering of documents
 - b. Marketing
 - i. Discussion boards analysis
 - ii. Blogs analysis
 - iii. Creation of perceptual maps
 - c. Accounting
 - i. Analysis of SEC filings (10Ks, 10Qs, 8Ks)
 - ii. Automatic Detection of Problematic Issues in company reports
 - d. News Analysis
 - i. Named Entity Extraction
 - ii. Event Detection
 - iii. Social Networks Analysis
 - iv. Trend Analysis
 - e. BioTech
 - i. Relations between genes, proteins, drugs, diseases
 - ii. Monitoring Company Drug Development Activities
 - f. Competitive Intelligence
 - i. Analyzing competitors press releases and web sites
 - g. Anti Terror Applications
 - i. 9/11 analysis
 - ii. Connectivity Analysis
 - iii. Centrality Analysis
 - iv. Blocks Modeling
 - 10. Text Mining Packages
 - a. Stanford NLP tools
 - b. Analyst Notebook
 - c. NetMap

חומר חובה לקריאה:

The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data (Hardcover) by Ronen Feldman and James Sanger, Cambridge University Press

The Handbook of Data Mining (N. Ye, ed.). Lawrence-Erlbaum Associates.

Handbook of Data Mining and Knowledge Discovery Edited by WILLI KLÖSGEN, Fraunhofer Institute for Autonomous Intelligent Systems, Sankt Augustin, Germany, and the late JAN M. ZYTKOW

KDD-2000-2014 Conference on Knowledge Discovery and Data Mining, proceedings and CD-ROM, ACM Press

George Chang, Marcus J. Healey, James A. M. McHugh, Jason T. L. Wang, Mining the World Wide Web: An Information Search Approach, Kluwer Academic Publishers, 2001, ISBN 0-7923-7349-9

-R. Kohavi, M. Spiliopoulou, J. Srivastava, editors, WEBKDD'2000 Web Mining for E Commerce -- Challenges and Opportunities, KDD-2000 workshop proceedings, August 2000, Boston, MA

R Feldman Techniques and applications for sentiment analysis Communications of the ACM 56 (4), 82-89

B Rozenfeld, R Feldman Self-supervised relation extraction from the Web Knowledge and Information Systems 17 (1), 17-33

-O Netzer, R Feldman, J Goldenberg, M Fresko Mine your own business: Market structure surveillance through text mining Marketing Science 31 (3), 521-543

חומר לקריאה נוספת:

מרכיבי הציון הסופי :

הגשת עבודה מסכמת / פרויקט גמר / מטלת סיכום / מבחן בית / רפרט 70 %
מטלות הגשה במהלך הסמסטר: תרגילים / עבודות / מבדקים / דוחות / פורום / סימולציה ואחרות
10 %
מצגת / הצגת פוסטר / הרצאה 20 %

מידע נוסף / הערות:

