

The Hebrew University of Jerusalem

Syllabus

TEXT MINING FOR BUSINESS APPLICATIONS - 55703

Last update 08-09-2018

HU Credits: 3

Degree/Cycle: 2nd degree (Master)

Responsible Department: Business Administration

Academic year: 0

Semester: 1st Semester

Teaching Languages: English

Campus: Mt. Scopus

Course/Module Coordinator: Prof. Ronen Feldman

Coordinator Email: ronen.feldman@huji.ac.il

Coordinator Office Hours: Monday 2pm-3pm

Teaching Staff:

Prof Ronen Feldman

Course/Module description:

The course provides an overview of the main techniques and applications of the text-mining field. Main topics are information categorization, information extraction, building crawlers for data gathering and sentiment analysis.

Course/Module aims:

The objective of the course is to provide methods for text-mining and experiment with building systems for analyzing large collections of documents. The course focuses on different techniques, algorithms for realizing these techniques and applications of the techniques in the business world.

Learning outcomes - On successful completion of this module, students should be able to:

Design and build a fundamental text mining system for analyzing large collections of documents. Students will be able to apply various methods for information categorization, information extraction, building crawlers for data gathering and sentiment analysis.

We will learn how to use R for building Text Mining Systems

Attendance requirements(%):

70%

Teaching arrangement and method of instruction: Frontal lecture of the material. Students' lectures on using text-mining techniques for solving a business problem. Execution of applied research in teams of 2-3 students.

Course/Module Content:

- 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

Required Reading:

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

Additional Reading Material:

Course/Module evaluation:

End of year written/oral examination 20 %

Presentation 10 %

Participation in Tutorials 0 %

Project work 0 %

Assignments 0 %

Reports 0 %

Research project 70 %

Quizzes 0 %

Other 0 %

Additional information: