



PhD Course: Text Analysis and Knowledge Mining

Course Outline

Faculty:	Education and Psychology
Departement:	Knowledge and Information Science
Credit hours:	3
Lecture time:	Tuesday 13-14
Laboratory time:	Tuesday 8-10
Instructor:	Dr. Mehrdad CheshmehSohrabi

Course Schedule

This course begins on September 24, 2024 and ends on December 31, 2024.

Duration:	16 weeks, 48 hours in total (theoretical topics, 16 hours; practical topics and activities, 32 hours)
Lectures:	16 hours , 1 hour per week
Laboratory:	32 hours, 2 hours per week, on project assignment)
Seminars:	3 hours (per week)

Course Objective

- To introduce students to the basic concepts and techniques of data mining, text analysis, and knowledge mining
- To familiarize students with various data and text mining applications with an emphasis on libraries
- To familiarize students with significant data and text mining tasks and algorithms
- To develop skills in data and text mining using RapidMiner/Python software

Syllabus

WEEK	TOPIC	REFERENCE
Week 1	Course overview: Why data mining, text mining and knowledge mining? applications of data & text mining	R1, R2 & R3
Week 2	Data warehouse	R1 & R2
Week 3	Data preprocessing	R2: Chapter 3
Week 4	Descriptive data mining	R1&R2
Week 5	Predictive data mining	R1&R2
Week 6	Data mining in library and information centers	<u>Slide</u>
Week 7	MIDTERM EXAM	-
Week 8	Text categorization: Naive Bayes, K Nearest Neighbor (KNN), Logistic Regression, Support Vector Machines, and Decision Trees.	R2: Chapter 10 & 11
Week 9	Text clustering	R3: Chapter 4
Week 10	Topic modeling: Probabilistic Latent Semantic Indexing (pLSI) and Latent Dirichlet Allocation (LDA)	<u>R3</u> : Chapter 5 & 8
Week 11	Text summarization	R3: Chapter 3
Week 12	Social media and network analysis	<u>R3</u> : Chapter 12
Week 13	Sentiment analysis	<u>R3</u> : Chapter 13, <u>Slide</u>
Week 14	Text visualization	<u>R5</u> : Chapter 10
Week 15	Presentation of projects	-
Week 16	FINAL EXAM	-

Outlier detection, Time series mining

Textbook

- R 1:** Tan, P. N., Steinbach, M., & Kumar, V. (2018). *Introduction to Data Mining*. Pearson Australia Pty Limited.
- R 2:** Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Waltham: Morgan Kaufmann.
- R 3:** Aggarwal, C. C., & Aggarwal, C. C. (2015). *Mining text data*. Springer International Publishing.

R4:

Witten, I. H., & Frank, E. (2002). *Data mining: practical machine learning tools and techniques with Java implementations*. Acm Sigmod Record, 31(1), 76-77.

R5:

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.

Reading List

- Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39, pp. 234-265). Cambridge: Cambridge University Press.
- Du, Y., & Khan, H. R. (2020). *Data science for librarians*. ABC-CLIO.
- Lamba, M., & Madhusudhan, M. (2022). *Text Mining for Information Professionals*. Springer.
- Bajpai, R. P., & Dwivedi, R. K. (2004). *Use of data mining in the field of library and information science: an overview*.
- You, S., Joo, S., & Katsurai, M. (2022). *Data mining topics in the discipline of library and information science: analysis of influential terms and Dirichlet multinomial regression topic model*. Aslib Journal of Information Management, (ahead-of-print).
- Sengupta, S. (2017). *Applications of data mining in library & information centres: an overview*. International Journal of Current Research, 9(01), 45246-45249

Learning Outcomes

On completion of this course, the student will be able to:

- Describe applications of data mining, text analysis and knowledge mining
- Describe and apply important data and text mining techniques and algorithms in data and text mining projects
- Perform pre-processing and text mining analysis in Rapidminer software
- Perform pre-processing and data and text mining analysis in Python software
- Perform some data mining projects

Software/Tool

- Assignment 1: Learning to use RapidMiner
- Assignment 2: Learning to use WEKA
- Assignment 2: Learning to use Python

Project

- Data mining in library

Class participation	10%
Homeworks	30%
Midterm Exam	20%
Final Exam	20%
Final Project	20%

Contacts

Dr. Mehrdad CheshmehSohrabi

Tel: 00983137935547

Email: mo.sohrabi@edu.ui.ac.ir