



## Course Syllabus

<b>Course Code</b>	<b>Course Title</b>	<b>ECTS Credits</b>
COMP-547DL	Social and Web Data Mining	10
<b>Prerequisites</b>	<b>Department</b>	<b>Semester</b>
COMP-544DL	Computer Science	Fall
<b>Type of Course</b>	<b>Field</b>	<b>Language of Instruction</b>
Elective	Data Science	English
<b>Level of Course</b>	<b>Lecturer(s)</b>	<b>Year of Study</b>
2 <sup>nd</sup> Cycle	Dr Ioannis Katakis	1 <sup>st</sup>
<b>Mode of Delivery</b>	<b>Work Placement</b>	<b>Corequisites</b>
Distant Learning	N/A	None

### Course Objectives:

The main objectives of the course are to:

- Explain the connection of information retrieval with search engines and how they work.
- Present the main elements of Link Analysis and more specifically to introduce algorithms like PageRank and HITS.
- Provide examples of how sentiment analysis can be applied on multiple social and business applications
- Define the main types of recommendation systems (item-based, user-based).
- Demonstrate how graph mining can be applied on social networks and provide examples of problems that can be solved with these techniques.
- Present advanced methods for web crawling and structured data extraction (wrappers).
- Explain the challenges of extracting knowledge from educational data.

### Learning Outcomes:

After completion of the course students are expected to be able to:

1. Explain the basic concepts of information retrieval.
2. Apply the HITS and PageRank algorithms on an artificial dataset.
3. Present the different applications of sentiment analysis and opinion mining.
4. Know how the analysis of user-generated data in web server logs can lead to improving web sites and understanding the users.

5. Explain how recommendation systems work and give examples of real world applications.
6. Apply graph mining algorithms to social networks.
7. Know how to identify events through the analysis of social networks.
8. Implement a personal search engine.
9. Can provide the pseudocode of the PageRank and HITS algorithms.
10. Discuss the potential application of data science in the education process

**Course Content:**

1. Information Retrieval and Web Search
  - a. Basic Concepts of Information Retrieval
  - b. Relevance Feedback
  - c. Evaluation Measures
  - d. Text and Web Page Pre Processing
  - e. Inverted Index
  - f. Latent Semantic Indexing
2. Link Analysis
  - a. PageRank
  - b. HITS
3. Web Crawling
4. Structured Data Extraction
5. Sentiment Analysis and Opinion Mining
  - a. Applications
  - b. The problem of Sentiment Analysis
  - c. Document Sentiment Classification
  - d. Sentence Subjectivity and Sentiment Classification
6. Web Usage Mining
  - a. Data Collection and Pre-Processing
  - b. Data Modeling
7. Recommendation Systems
  - a. Basic Concepts (Utility Matrix, Long Tail, Applications)
  - b. Content-Based Recommendations
  - c. Collaborative Filtering
  - d. Dimensionality Reduction
8. Mining Social-Network Graphs
  - a. Social Networks as Graphs
  - b. Clustering of Social Network Graphs
  - c. Community Detection
  - d. Event Detection in Social Networks
9. Educational Data Mining and Learning Analytics

- a. What is Educational Data Mining
- b. Sources of Educational Data
- c. Algorithms and Applications of Learning Analytics

**Learning Activities and Teaching Methods:**

Lectures, Demonstration of Web Mining and Sentiment Analysis Tools, Assignments, Projects.

**Assessment Methods:**

Projects

Final Assessment\*

\* The Final Assessment can be either a Final Exam or Final Assignment(s) with Viva Participation/Homework Assignments/Quizzes

**Required Textbooks / Readings:**

Title	Author(s)	Publisher	Year	ISBN
<i>Mining Massive Datasets (2<sup>nd</sup> Edition)*</i>	Jure Leskovec, Anand Rajaraman, Jeff Ullman	Cambridge University Press	2015	978-1107077232 (Free e-book)
<i>Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data **</i>	Bing Liu	Springer	2011	978-3642194597
<i>Data Mining – The Textbook</i>	Charu C. Aggarwal	Springer	2015	978-3-319-14141-1
<i>Educational Data Mining and Learning Analytics (chapter in Learning Analytics)</i>	Baker R.S., Inventado P.S.	Springer	2014	978-1-4614-3304-0

\* Freely available

\*\* Freely available through the University of Nicosia network

**Recommended Textbooks / Readings:**

<b>Title</b>	<b>Author(s)</b>	<b>Publisher</b>	<b>Year</b>	<b>ISBN</b>
Networks, Crowds, and Markets: Reasoning about a Highly Connected World	David Easley, Jon Kleinberg	Cambridge University Press	2010	978-0521195331 <i>(Free e-book)</i>
Sentiment Analysis: Mining Opinions, Sentiments, and Emotions	Bing Liu	Cambridge University Press	2015	978-1107017894
Machine Learning for the Web	Andrea Isoni	PACKT Publishing	2016	978-1-78588-660-7