# COURSE OUTLINE

## GENERAL

| | | | |
|---|---|---|---|
| **SCHOOL** | Sciences and Engineering | | |
| **ACADEMIC UNIT** | Computer Science | | |
| **LEVEL OF STUDIES** | 1st Cycle | | |
| **COURSE CODE** | COMP-446 | **SEMESTER** | Fall |
| **COURSE TITLE** | Web and Social Data Mining | | |

| **INDEPENDENT TEACHING ACTIVITIES** *if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits* | **WEEKLY TEACHING HOURS** | **CREDITS** |
|---|---|---|
| | 2.5 | 6 |
| | | |
| | | |
| *Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).* | | |

| | |
|---|---|
| **COURSE TYPE** *general background, special background, specialised general knowledge, skills development* | Specialization |
| **PREREQUISITE COURSES:** | COMP-244 |
| **LANGUAGE OF INSTRUCTION and EXAMINATIONS:** | English |
| **IS THE COURSE OFFERED TO ERASMUS STUDENTS** | |
| **COURSE WEBSITE (URL)** | |

## LEARNING OUTCOMES

**Learning outcomes**

*The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.*

*Consult Appendix A*

- *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area*
- *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B*
- *Guidelines for writing Learning Outcomes*

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

- Explain and analyze the advanced theoretical foundations and basic concepts of information retrieval.
- Apply and evaluate the HITS and PageRank algorithms on complex, artificial datasets, interpreting their implications.
- Present, analyze, and differentiate the diverse applications of sentiment analysis and opinion mining, considering their practical and ethical dimensions.
- Demonstrate an in-depth understanding of how the analysis of user-generated data in web server

logs can lead to strategic improvement of websites and a comprehensive understanding of user behavior.
- Explain, evaluate, and provide examples of how recommendation systems function, illustrating their real-world applications and underlying mechanisms.
- Apply and adapt advanced graph mining algorithms to social networks, deriving meaningful insights and identifying potential challenges.
- Design, develop, and implement a personal search engine, demonstrating mastery of the underlying principles and technologies.
- Formulate and present the pseudocode of the PageRank and HITS algorithms, along with a critical explanation of their operational details and limitations.

**General Competences**

*Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?*

| | |
|---|---|
| *Search for, analysis and synthesis of data and information, with the use of the necessary technology* | *Project planning and management* |
| *Adapting to new situations* | *Respect for difference and multiculturalism* |
| *Decision-making* | *Respect for the natural environment* |
| *Working independently* | *Showing social, professional and ethical responsibility and sensitivity to gender issues* |
| *Team work* | *Criticism and self-criticism* |
| *Working in an international environment* | *Production of free, creative and inductive thinking* |
| *Working in an interdisciplinary environment* | *……* |
| *Production of new research ideas* | *Others…* |
| | *…….* |

Search for, analysis and synthesis of data and information, with the use of the necessary technology
Adapting to new situations
Decision-making
Working independently
Production of new research ideas
Project planning and management
Criticism and self-criticism
Production of free, creative and inductive thinking
Working in an interdisciplinary environment

**SYLLABUS**

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.Sentiment Analysis and Opinion Mining
   a. Applications

b. The problem of Sentiment Analysis
c. Document Sentiment Classification
d. Sentence Subjectivity and Sentiment Classification

4.Web Usage Mining
a. Data Collection and Pre-Processing
b. Data Modeling

5.Recommendation Systems
a. Basic Concepts (Utility Matrix, Long Tail, Applications)
b. Content-Based Recommendations
c. Collaborative Filtering
d. Dimensionality Reduction

6.Mining Social-Network Graphs
a. Social Networks as Graphs
b. Clustering of Social Network Graphs
c. Community Detection
d. Event Detection in Social Networks

**TEACHING and LEARNING METHODS - EVALUATION**

| | |
|---|---|
| **DELIVERY**<br>*Face-to-face, Distance learning, etc.* | Face-to-face |
| **USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY**<br>*Use of ICT in teaching, laboratory education, communication with students* | *Use of ICT in teaching / Χρήση ΤΠΕ*<br>*Communication with students / Επικοινωνία με Φοιτητές* |
| **TEACHING METHODS**<br>*The manner and methods of teaching are described in detail.*<br>*Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.*<br><br>*The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS* | |

| ***Activity*** | ***Semester workload*** |
|---|---|
| Lectures | 35 |
| Preparation, Homework, Quizzes | 77 |
| Exam Preparation | 36 |
| Final Exam | 2 |
| | |
| Course total | ***150*** |

| | |
|---|---|
| **STUDENT PERFORMANCE EVALUATION**<br>*Description of the evaluation procedure*<br><br>*Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other* | - Mid-term exam<br>- Final Examination<br>- Participation/Homework Assignments/Quizzes |

| | |
|---|---|
| *Specifically-defined evaluation criteria are given, and if and where they are accessible to students.* | |

**ATTACHED BIBLIOGRAPHY**

**Required Textbooks / Readings**

| Authors | Title | Publisher | Year | ISBN |
|---|---|---|---|---|
| Jure Leskovec, Anand Rajaraman, Jeff Ullman | *Mining Massive Datasets (2nd Edition)* | Cambridge University Press | 2014 | 978-1107077232 (Free e-book) |
| Bing Liu | *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data* | Springer | 2011 | 978-3642194597 |

**Recommended Textbooks / Readings:**

| Authors | Title | Publisher | Year | ISBN |
|---|---|---|---|---|
| David Easley, Jon Kleinberg | Networks, Crowds, and Markets: Reasoning about a Highly Connected World | Cambridge University Press | 2010 | 978-0521195331 (Free e-book) |
| Bing Liu | Sentiment Analysis: Mining Opinions, Sentiments, and Emotions | Cambridge University Press | 2015 | 978-1107017894 |