UNIVERSITAT D'ANDORRA

Seminar's Teaching Plan

Teaching plan	Bachelor in Computer Science

Seminar	Database Design and Data Processing
Semester	4
Coordination **Professor A/e**	Jan Sau Batlle jsau@uda.ad
Mode	Presential and virtual
Teaching language	English

1. Seminar summary

Like any other engineering, you need to be able to track a project. This involves the implementations of a series of activities, such as the application of a methodology for planning, organizing and developing a project, combined with the analysis of data in order to specify and design a software system and, in special, extract information relevant to organizations.

In this seminar we will look at the different types of existing databases and what can be done with the data stored in them.

2. Seminar Content

- **Non-relational databases**
  - Introduction
  - Differences with the relational model
  - Key-Value Storage
  - Column storage
  - Document stores
  - Graph models
  - Combined models

- Spatial databases (GIS)
- Common tools and their Applications

- **Data mining**
  - Introduction and objectives in data mining
  - The iterative learning process
  - Questions the company asks
  - Understanding the data
  - Data filtering, cleaning and combining
  - Getting new information
  - Data modeling
  - Data visualization
  - Validation of results
  - Prediction
  - The workflow using computer tools

- **Big Data**
  - What is and what is not big data
  - The V of the massive data
  - Lots of real-time and varied information
  - Storage or treatment
  - The MapReduce algorithm
  - Introduction to the Hadoop ecosystem
  - A tool for every need
  - Distributed file systems
  - Distributed databases
  - Indexing and search
  - Machine learning about massive data
  - Introduction to Deep Learning
  - Tools for mass data processing

3. Seminar Activities

3.1. Continuous Evaluation

The continuous assessment includes the performance of different assessment activities during the semester:

- 3 Virtual Projects (TV)
- 3 In person tests (CP)
- Modul's Challenge

With the following weighting of learning outcomes:

	CP 1	CP 2	CP 3	TV 1	TV 2	TV 3	CHALLENGE Mandatory	Total Evaluation
BInfo-E005-04- Understands the particularities of database management systems beyond relational models.	20 %	20 %	20 %	0 %	0 %	20 %	20 %	**100 %**
BInfo-E005-05- Applies machine learning algorithms on heterogeneous data sets to extract information relevant to organizations.	0 %	0 %	0 %	30 %	50 %	0 %	20 %	**100 %**
BInfo-E005-06- Knows the ecosystem of mass data processing and knows how to apply workflows on them.	0 %	0 %	0 %	40 %	20 %	20 %	20 %	**100 %**

3.2. Final Evaluation

If you opt for the **Final Exam (EF)** evaluation, the weighting of the learning results of the final assessment will be:

	EF	TV 1	TV 2	TV 3	CHALLENGE Mandatory	Total Evaluation
BInfo-E005-04- Understands the particularities of database management systems beyond relational models.	60 %	0 %	0 %	20 %	20 %	**100 %**
BInfo-E005-05- Applies machine learning algorithms on heterogeneous data sets to extract information relevant to organizations.	0 %	30 %	50 %	0 %	20 %	**100 %**
BInfo-E005-06- Knows the ecosystem of mass data processing and knows how to apply workflows on them.	0 %	40 %	20 %	20 %	20 %	**100 %**

IMPORTANT:

In both assessments (continuous and final exam) it is mandatory to submit and present the challenge in order to pass the Module.

4. Seminar Resources

Basic Bibliography

Professor material

Professor prompts.

Moodle material.

Reference books

- L. Igual and S. Seguí, "Introduction to Data Science", Springer, 2016.
- I. D. Dinov, "Data Sciene and Predictive Analytics", Springer, 2018.
- M. Bramer, "Principles of Data Mining", Springer, 2016.

Other ressources
• Excel • Python • BeautifullSoup4

5. Comments

- No exam (CP or ExF) will be repeated.
- Late submissions will not be accepted.
- The continuous assessment (CP) exams are face-to-face tests and will be set in the seminar schedule.
- The final exam (ExF) and the recovery of the challenge are face to face test and are carried out on the date indicated in the seminar recovery calendar published by the UdA.
- If a student does not submit a TR assignment or doesn't attend an ExAC exam, the grade for that assignment or exam is zero.
- All evaluable submissions will be made on the UdA campus and in the format indicated in the statement of work or challenge (the UdA campus does not accept submissions larger than 50 MB).