Course guide
270951 - ADSDB - Algorithms, Data Structures and Databases

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.

Degree: MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Compulsory subject).

Academic year: 2022  ECTS Credits: 6.0  Languages: English

LECTURER

Coordinating lecturer: MARIA JOSEFINA SIERRA SANTIBAÑEZ - ANNA QUERALT CALAFAT

Others: Primer quadrimestre:
ANNA QUERALT CALAFAT - 11, 12
OSCAR ROMERO MORAL - 11, 12
MARIA JOSEFINA SIERRA SANTIBAÑEZ - 11, 12

PRIOR SKILLS

This course assumes basic competences in algorithms, data structures and databases. The course is structured to cope with different backgrounds and learning needs but basic knowledge on Computer Science principles is assumed: notions of computer architecture, basic programming constructs and data structures.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CE2. Apply the fundamentals of data management and processing to a data science problem

General:
CG1. Identify and apply the most appropriate data management methods and processes to manage the data life cycle, considering both structured and unstructured data

Transversal:
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

Basic:
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
TEACHING METHODOLOGY

The students are divided in two tracks: one for students with a minor in computer science (track 1) and another one for students with a major in computer science (track 2).

Students in track 1 will study fundamental concepts in algorithms, data structures and databases. First, additional material to read, study and understand is provided. Lectures focus on the main concepts and those that require some additional explanation to guarantee a proper understanding. Students will have a large bank of self-assessing exercises to practice their understanding on their own. During the course, they will have to solve some mandatory exercises to guarantee a smooth learning process. Additionally, in the face-to-face lectures, the lecturer will solve doubts, go through representative exercises to guarantee a solid understanding and discuss exercises (to be solved during the lecture) with the students.

Students in track 2 will carry out the DS-EtE project and, for them, this course is a project course. There, students must create an end-to-end system architecture to ingest, store, process, learn models and deploy such system for a realistic project with realistic data. The students must develop good practices developing such architecture (what nowadays is known as DataOps / MLOps).

This course has a strong self-learning component and shares the same objectives in both tracks. The lecturers will supervise the students progress during the semester to guarantee a proper progress.

LEARNING OBJECTIVES OF THE SUBJECT

1. To analyse the cost of iterative and recursive algorithms
2. To review some simple data structures: stacks, queues, lists, and trees
3. To know, explain, design, analyse, compare and implement the main data structures and algorithms that can be used to implement priority queues
4. To know, explain, design, analyse, compare and implement the main data structures and algorithms that can be used to implement dictionaries
5. To know, explain, design, analyse, compare and implement the main data structures and algorithms that can be used to represent graphs and solve classic graph problems such as traversals, topological ordering and shortest paths
6. To know, understand, explain, analyse and compare some algorithm design techniques: greedy, divide and conquer, and dynamic programming
7. To be aware of the limits of computation: to understand the definitions of the P and NP classes, the concept of Polynomial-Time reduction, the notion of NP-Completeness, and to know some classic NP-complete problems
8. Describe what is a database and a database management system
9. Effectively use the standard Structured Query Language (SQL) to query relational databases
10. Explain the relational data model, including its data structures, the relational algebra and integrity constraints
11. Given a set of informational requirements, model the logic schema of a relational database
12. Identify the main objectives of a database management system query optimizer
13. Develop a quality realistic end-to-end system architecture for a data science project

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self study</td>
<td>96,0</td>
<td>64.00</td>
</tr>
<tr>
<td>Hours small group</td>
<td>24,0</td>
<td>16.00</td>
</tr>
<tr>
<td>Guided activities</td>
<td>6,0</td>
<td>4.00</td>
</tr>
<tr>
<td>Hours large group</td>
<td>24,0</td>
<td>16.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h
## CONTENTS

<table>
<thead>
<tr>
<th>Basics of Analysis of Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simple Data Structures: Review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Stacks, queues, lists and trees.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Priority Queues</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dictionaries</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Graphs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm design techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Greedy, divide and conquer, and dynamic programming.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Introduction to NP and Computational Intractability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Basic introduction to P and NP classes, Polynomial-Time reduction, and NP-completeness. Examples of classic NP-complete problems.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Introduction to databases and database management systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
</tr>
<tr>
<td>Main concepts on databases and database management systems. Relational database management systems.</td>
</tr>
<tr>
<td>Subject</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>SQL: Data-definition language and data-manipulation language</td>
</tr>
<tr>
<td>The relational model</td>
</tr>
<tr>
<td>The relational algebra</td>
</tr>
<tr>
<td>Logical design of relational databases</td>
</tr>
<tr>
<td>Notions of physical design and physical database optimization</td>
</tr>
<tr>
<td>Data Science-related advanced topics</td>
</tr>
</tbody>
</table>
ACTIVITIES

Basics of Analysis of Algorithms

Description:
To compare the efficiency of different algorithms for solving the same problem and select the most appropriate one. To compute the cost of an algorithm in the worst, best and average cases. To understand the definitions of the asymptotic order of growth notations Big-O, Omega and Theta, and their usefulness in characterising algorithm efficiency in time and space.

Specific objectives:
1

Related competencies:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 5h
Theory classes: 3h
Practical classes: 2h

Review of simple data structures

Description:
Operations. Lists, Stacks, Queues, Trees

Specific objectives:
2

Related competencies:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 4h
Theory classes: 1h
Laboratory classes: 3h
Priority Queues

Description:

Specific objectives:
3

Related competencies:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 4h
Theory classes: 1h
Laboratory classes: 3h

Dictionaries

Description:
Operations of dictionaries and ordered dictionaries. Basic implementations: tables and lists. Advanced implementations: hash tables, binary search trees, AVL trees

Specific objectives:
4

Related competencies:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 4h
Theory classes: 1h
Laboratory classes: 3h
Graphs

Description:

Specific objectives:
5

Related competencies:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 6h
Theory classes: 2h
Laboratory classes: 4h

Algorithmic schemes

Description:
Divide and conquer, Greedy algorithms, Dynamic Programming, Exhaustive search, Backtracking.

Specific objectives:
6

Related competencies:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Notions of Intractability

Description:
Basic introduction to P and NP classes. NP-completeness.

Specific objectives:
7

Related competencies:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
Partial Exam

Description:
For students with a minor in Computer Science, this exam evaluates their knowledge on fundamental concepts of algorithms, data structures and databases

Specific objectives:
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Related competencies:
CE2. Apply the fundamentals of data management and processing to a data science problem
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 8h
Guided activities: 2h
Self study: 6h

Introduction to databases and database management systems

Description:
The student attends the lecture, takes notes and participates in the session exercises

Specific objectives:
8

Related competencies:
CE2. Apply the fundamentals of data management and processing to a data science problem
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 1h
Theory classes: 1h
**SQL**

**Description:**
The student attends the lecture, takes notes and participates in the session exercises

**Specific objectives:**
9

**Related competencies:**
CE2. Apply the fundamentals of data management and processing to a data science problem
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

**Full-or-part-time:** 3h
Laboratory classes: 3h

**The Relational Model**

**Description:**
The student attends the lecture, takes notes and participates in the session exercises

**Specific objectives:**
10, 11

**Related competencies:**
CE2. Apply the fundamentals of data management and processing to a data science problem
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

**Full-or-part-time:** 3h
Theory classes: 1h
Laboratory classes: 2h
## Logical Design of Relational Databases

**Description:**
The student attends the lecture, takes notes and participates in the session exercises

**Specific objectives:**  
10, 11

**Related competencies:**  
CE2. Apply the fundamentals of data management and processing to a data science problem  
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

**Full-or-part-time:** 4h  
Theory classes: 1h  
Laboratory classes: 3h

## Physical Optimization

**Description:**
The student attends the lecture, takes notes and participates in the session exercises

**Specific objectives:**  
12

**Related competencies:**  
CE2. Apply the fundamentals of data management and processing to a data science problem  
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

**Full-or-part-time:** 7h  
Theory classes: 3h  
Laboratory classes: 4h
# Final Exam

**Description:**
For students with a minor in Computer Science, this exam evaluates their knowledge on fundamental concepts of algorithms, data structures and databases.

**Specific objectives:**
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

**Related competencies:**
- CE2. Apply the fundamentals of data management and processing to a data science problem
- CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
- CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

**Full-or-part-time:** 7h
- Guided activities: 2h
- Self study: 5h

---

# Data Science End-to-End Project (DS-EE)

**Description:**
All students will undertake a project spanning all main phases of a data science. As result, they are asked to develop a quality realistic end-to-end system architecture for a data science project.

**Specific objectives:**
13

**Related competencies:**
- CG1. Identify and apply the most appropriate data management methods and processes to manage the data life cycle, considering both structured and unstructured data
- CE2. Apply the fundamentals of data management and processing to a data science problem
- CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
- CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

**Full-or-part-time:** 92h
- Laboratory classes: 12h
- Self study: 80h
GRADING SYSTEM

Let DBM = Databases midterm exam grade,
DBF = Databases final exam grade,
DBEx = Database exercises to be solved during the course,
ADSM = Algorithms and Data Structures midterm exam grade,
ADSF = Algorithms and Data Structures final exam grade and
ADSEX = Algorithms and Data Structures exercises to be solved during the course,
DS-EtE = DS-EtE project grade

Then,

1) If the student followed track 1 (see methodology) the mark is calculated as follows:
BD = MAX (0.2*DBEx + 0.8*DBM, DBF),
ADS = MAX (0.2*ADSEX + 0.4*ADSM + 0.4*ADSF, ADSF).

ADSDB (final course mark) = 0.5*BD + 0.5*ADS

2) If the student followed track 2 (see methodology) the mark is calculated as follows:

ADSDB (final course mark) = DS-EtE

BIBLIOGRAPHY

Basic:

RESOURCES

Hyperlink: