UNIVERSIDAD POLITÉCNICA DE MADRID
FACULTAD DE INFORMÁTICA
Campus de Montegancedo
Boadilla del Monte. 28660 Madrid

Data Engineering

Learning Guide – Information for Students

1. Description

Grade	EUROPEAN MASTER ON SOFTWARE ENGINEERING
Module	Advanced Software Engineering Aspects
Area	
Subject	Data engineering
Type	Optional
ECTS credits	4
Responsible department	DLSIIS
Major/Section/	Unknown

Academic year	2012/2013
Term	1st or 3rd term
Language	English
Web site	http://pegaso.ls.fi.upm.es/idatos

2. Faculty

NAME and SURNAME	OFFICE	email
Óscar Marbán Gallego	D-4302	omarban@fi.upm.es
Ernestina Menasalvas Ruiz	D-4303	emenasalvas@fi.upm.es

3. Prior knowledge required to take the subject

Passed subjects	•
Other required learning outcomes	• Design of relational databases • Implementation of relational databases • SQL language

4. Learning goals

SUBJECT-SPECIFIC COMPETENCES AND PROFICIENCY LEVEL		
Code	**Competence**	**Level**
SC13	To have a vision of the different specific and emergent aspects of the Software Engineering, and to go further in some of them.	S
SC14	To understand what nowadays software engineering procedures can and cannot reach, their limitations and their possible future evolution.	S

Proficiency level: knowledge (K), comprehension (C), application (A), and analysis and synthesis (S)

SUBJECT LEARNING OUTCOMES			
Code	**Learning outcome**	**Related competences**	**Proficiency level**
LR1	Faced with a real problem, choose the most appropriate software engineering solution, analyzing the feasibility of this solution, what can and can not get through the current state of development of the chosen solution, and which is expected advances in the near future.	SC13, SC14	3-4

5. Subject assessment system

ACHIEVEMENT INDICATORS		
Ref	**Indicator**	**Related to LR**
I1	Design of a Data Warehouse	RA1
I2	Knowledge of the architecture of a Data Warehouse	RA1
I3	Implements a Data Warehouse	RA1
I4	Obtains implici knowlegde from data	RA1
I5	Knowlede of Data Mining techniques	RA1
I6	Knowledge of the Data Mining process	RA1
I7	Handle Data Mining techniques	RA1
I8	Handles Data Mining tools	RA1

CONTINUOUS ASSESSMENT			
Brief description of assessable activities	**Time**	**Place**	**Weight in grade**
Project 1 (4 assignements)	Weeks 1-8	Home	40%
Project 2	Weeks 8 - 16	Home	40%
Exam	Week 16	Class	20%
			Total: 100%

UNIVERSIDAD POLITÉCNICA DE MADRID
FACULTAD DE INFORMÁTICA
Campus de Montegancedo
Boadilla del Monte. 28660 Madrid

GRADING CRITERIA

The course is composed by two main parts:

1. Data Warehouse
2. Data Mining

Each part will be evaluated taking into account the evaluation of the final exam and a final Project as follows.

1. Part 1. Project 1 composed by 4 assignments that will have to be presented in the classroom. An exam will also take place and the values of it can be seen in the table.
2. Part 2. Proyect 2 that will have to be presented in the classroom when assigned. There will be also some assignments and a final exam. The values of each part can be seen in the following table.

Projects will be developed by working groups composed by 4 students enrolled in the course.

In order to pass the course the following requirements are needed:

1. to obtain 50/100 points at least
2. It is MANDATORY to do the exam and to complete all the assignments and projects
3. For each Project it is required to obtain at least a qualification of 30% from its final value (see table)
4. For each exam the qualification has to be over the 30% of the maximun value of the exam (see table)

For those who do not pass the course there will be an extra opportunity. To pass this the following requirements are needed:

1. There will be no projects and there will be an exam of each part of the course.
2. The values of each part are as follows:
    a. Part 1: 50%
    b. Part 2: 50%

GRADING CRITERIA

3. Obtain 50/100 points in the global value

6. Contents and learning activities

SPECIFIC CONTENTS		
Unit / Topic / Chapter	**Section**	**Related indicators**
Chapter 1: INTRODUCTION	1.1 Business Intelligence Introduction	I2, I6
Chapter 2: Data warehouse	2.1 Data warehouse architecture	I2
	2.2 ETL Tools	I1, I2
	2.3 Multidimensional model	I1, I3
	2.4 Data warehouse design methodologies	I1, I3
	2.5 Data warehouse tools	I1, I3
	2.6 OLAP tools	I4
Chapter 3: Data Mining	3.1 KDD process	I6
	3.2 Data mining techniques	I7
	3.3 Data mining tools	I8
	3.4 Data mining methodologies	I6
	3.5 CRISP-DM	I6

7. Brief description of organizational modalities and teaching methods

TEACHING ORGANIZATION		
Scenario	**Organizational Modality**	**Purpose**
	Theory Classes	***Talk to students***
	Seminars/Workshops	***Construct knowledge through student interaction and activity***
	Practical Classes	***Show students what to do***
	Placements	***Round out student training in a professional setting***
	Personal Tutoring	***Give students personalized attention***
	Group Work	***Get students to learn from each other***
	Independent Work	***Develop self-learning ability***

UNIVERSIDAD POLITÉCNICA DE MADRID
FACULTAD DE INFORMÁTICA
Campus de Montegancedo
Boadilla del Monte. 28660 Madrid

TEACHING METHODS		
	Method	**Purpose**
	Explanation/Lecture	***Transfer information and activate student cognitive processes***
	Case Studies	***Learning by analyzing real or simulated case studies***
	Exercises and Problem Solving	***Exercise, test and practice prior knowledge***
	Problem-Based Learning (PBL)	***Develop active learning through problem solving***
	Project-Oriented Learning (POL)	***Complete a problem-solving project applying acquired skills and knowledge***
	Cooperative Learning	***Develop active and meaningful learning through cooperation***
	Learning Contract	***Develop independent learning***

Known as explanation, this teaching method involves the "*presentation of a logically structured topic with the aim of providing information organized according to criteria suited for the purpose".* This methodology, also known as *lecture*, mainly focuses on the verbal exposition by the teacher of contents on the subject under study. The term *master class* is often used to refer to a special type of lecture taught by a professor on special occasions

Intensive and exhaustive analysis of a real fact, problem or event for the purpose of understanding, interpreting or solving the problem, generating hypotheses, comparing data, thinking, learning or diagnosis and, sometimes, training in possible alternative problem-solving procedures.

Situations where students are asked to develop the suitable or correct solutions by exercising routines, applying formulae or running algorithms, applying information processing procedures and interpreting the results. It is often used to supplement lectures.

Teaching and learning method whose starting point is a problem, designed by the teacher, that the student has to solve to develop a number of previously defined competences.

Teaching and learning method where have a set time to develop a project to solve a problem or perform a task by planning, designing and completing a series of activities. The whole thing is based on developing and applying what they have learned and making effective use of resources.

Interactive approach to the organization of classroom work where students are responsible for their own and their peers' learning as part of a co-responsibility strategy for achieving group goals and incentives.

This is both one of a number of methods for use and an overall teaching approach, or philosophy.

An agreement between the teacher and student on the achievement of learning outcomes through an independent work proposal, supervised by the teacher, and to be accomplished within a set period. The essential points of a learning contract are that it is a written agreement, stating required work and reward, requiring personal involvement and having a time frame for accomplishment.

BRIEF DESCRIPTION OF THE ORGANIZATIONAL MODALITIES AND TEACHING METHODS	
THEORY CLASSES	Theoretical classes will proceed participatively, with certain ones requiring a previous study by the student of some bibliography
PROBLEM-SOLVING CLASSES	Some classes will be dedicated to problem solving and modelling in the classroom to have a discussion on the issues modeled at the end of the classroom.
PRACTICAL WORK	
INDIVIDUAL WORK	Individual works will be announced in classroom for the course.
GROUP WORK	The main learning activity in the course will consist on a team project in teams of 4 students, applying practically the concepts discussed in the classroom to the project.
PERSONAL TUTORING	The classes dedicated to follow-up of team projects will serve a group tutoring, while other assignments will be tutored at the student request.

8. Teaching resources

TEACHING RESOURCES	
RECOMMENDED READING	Building the data warehouse. W. H. immnon. 1996. Willey
	Managing the Data Warehouse. W. H. Immon. 1997. Willey
	Building the operational Data Store. W. H. Immon. 1999. Willey
	Exploration Datawarehouse. W. Immon. 2000. Willey
	Improving Data Warehouse and Business Information Quality. Methods for reducing cost and increasing profits. L. English. 1999 Willey
	Data Mining Techniques for Marketing, Sales, and Customer support. Michael J. A. Berry and Gordon Linoff. 1997. Willey
	Data Mining Solutions: Methods for solving Real-World Problems. C. Westphal, T. Blaxton. 1998. Willey
	Mastering Data mining . The art and science of Customer relationship management. M. Berry, G. Linoff. 2000. Willey
	The data warehouse lifecycle toolkit. R. Kimball. 2000. Willey
WEB RESOURCES	Subject web site (http://pegaso.ls.fi.upm.es)
EQUIPMENT	Room
	Group work room

9. Subject schedule

Week	Classroom activities	Lab activities	Individual work	Group work	Assessment activities	Others
Week 1 (6 hours)	• 1.1 Business Intelligence introduction (2 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	•	•
Week 2 (7 hours)	• 2.1 Data warehouse architecture (2 hours)	•	• Independent work (2 hours)	• Group work (3 hours)	•	•
Week 3 (7 hours)	• 2.2 ETL Tools (1 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	• Presentation (1 hours)	•
Week 4 (7 hours)	• 2.3 Multidimensional model (2 hours)	•	• Independent work (2 hours)	• Group work (3 hours)	•	•
Week 4 (7 hours)	• 2.3 Multidimensional model (2 hours)	•	• Independent work (2 hours)	• Group work (3 hours)	•	•
Week 6 (7 hours)	• 2.5 Data warehouse tools (1 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	• Presentation (1 hora)	•
Week 7 (7 hours)	• 2.6 OLAP Tools (1 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	• Presentation (1 hora)	•
Week 8 (7 hours)	• 2.4 Data warehouse design methodologies (1 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	• Presentation (1 hora)	•

Week 9 (6 hours)	• 3.1 KDD process (2 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	•	•
Week 10 (7 hours)	• 3.1 KDD process (2 hours)	•	• Independent work (2 hours)	• Group work (3 hours)	•	•
Week 11 (7 hours)	• 3.2 Data mining techniques (2 hours)	•	• Independent work (2 hours)	• Group work (3 hours)	•	•
Week 12 (7 hours)	• 3.2 Data mining techniques (1 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	• Presentation (1 hora)	•
Week 13 (7 hours)	• 3.3 Data mining tools (1 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	• Presentation (1 hora)	•
Week 14 (7 hours)	• 3.4 Data mining methodologies (2 hours)	•	• Independent work (2 hours)	• Group work (3 hours)	•	•
Week 15 (6 hours)	• 3.5 CRISP-DM (2 hours)	•	• Independent work (2 hours)	• Group work (2 hours)	•	•
Week 16 (7 hours)	• 3.5 CRISP-DM (1 hours)	•	• Independent work (2 hours)	• Group work (3 hours)	• Presentation (1 hora)	•
Week 17 (2 hours)	•	•	•	•	• Final exam (2 hours)	•
	•	•	•	•	•	•

Note: Student workload specified for each activity in hours