ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000899 - Open Data And Knowledge Graphs

DEGREE PROGRAMME

10BA - Master Universitario en Ciencia de Datos

ACADEMIC YEAR & SEMESTER

2020/21 - Semester 1

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	103000899 - Open Data And Knowledge Graphs
No of credits	4.5 ECTS
Type	Compulsory
Academic year ot the programme	First year
Semester of tuition	Semester 1
Tuition period	September-January
Tuition languages	English
Degree programme	10BA - Master Universitario en Ciencia de Datos
Centre	10 - Escuela Tecnica Superior de Ingenieros Informaticos
Academic year	2020-21

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Oscar Corcho Garcia (Subject coordinator)	2209	oscar.corcho@upm.es	Sin horario.
Raul Garcia Castro	2110	r.garcia@upm.es	Sin horario.

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

The subject - recommended (passed), are not defined.

3.2. Other recommended learning outcomes

- Basic knowledge about Web technologies

- Basic knowledge about database modelling

- Basic knowledge about knowledge representation techniques

4. Skills and learning outcomes *

4.1. Skills to be learned

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB09 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CECD01 - Conocer los procesos de captura, extracción, manipulación y conversión de datos en diferentes entornos.

CG14 - Capacidad de trabajar y comunicarse también en contextos internacionales

CGI01 - Adquirir conocimientos científicos avanzados del campo de la informática que le permitan generar nuevas ideas dentro de una línea de investigación

4.2. Learning outcomes

RA9 - Develop applications that exploit the (open) data available on Knowledge Graphs

RA5 - Identify and solve those types of real-world problems in which Open Data and Knowledge Graph technologies can be successfully applied

RA6 - Use different languages, techniques, methods and methodologies that enable the development of ontologies and data for Knowledge Graphs

RA7 - Develop ontologies that serve as vocabularies for the data available on Knowledge Graphs

RA8 - Generate data in the format used for Open Data and Knowledge Graphs, and to publish them for the use of third parties

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

During this course you will learn about the main foundations for Knowledge Graphs, including the W3C recommendations that are applicable in this area (RDF, RDF Schema, SPARQL, OWL, CSV on the Web) and methodologies for the generation and publication of Knowledge Graphs from multiple heterogeneous data sources and formats, using Linked Data technologies. You will also learn about how to create shared agreed vocabularies and ontologies that can give support to Knowledge Graph generation and publication, and will understand how these principles and approaches have been applied to several domains. You will work on a practical hands-on exercise that will be the basis for your participation in hackatons and other similar events.

5.2. Syllabus

1. Introduction to Open Data and Knowledge Graphs
2. Data and knowledge representation and access for Open Data and Knowledge Graphs
   - 2.1. Basic Overview on Knowledge Representation
   - 2.2. RDF and RDF Schema
   - 2.3. SPARQL
3. Knowledge Graph generation, linking and publication from heterogeneous data sources
   - 3.1. Methodological guidelines for Knowledge Graph generation
   - 3.2. RDF generation from relational databases
   - 3.3. RDF generation from semi-structured data sources
   - 3.4. RDF generation using declarative mappings
   - 3.5. Data linking
   - 3.6. Knowledge Graph publication as Linked Data
4. Vocabulary selection and development for Knowledge Graphs
   - 4.1. Methodologies for building vocabularies
   - 4.2. Existing vocabularies
5. Knowledge Graph-based applications and advanced research topics
   - 5.1. Knowledge Graph application architecture
   - 5.2. Advanced research topics in Open Data and Knowledge Graphs

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1			**Unit 1** Duration: 00:30 **Unit 2** Duration: 03:30	
2			**Unit 2** Duration: 00:30 **Unit 2** Duration: 03:30	
3			**Unit 2** Duration: 00:30 **Units 2 and 3** Duration: 03:30	**SPARQL test** Continuous assessment Not Presential Duration: 01:00
4			**Unit 3** Duration: 00:30 **Unit 3** Duration: 03:30	
5			**Unit 3** Duration: 00:30 **Unit 3** Duration: 01:30	
6			**Unit 3** Duration: 00:30 **Unit 3** Duration: 03:30	**RDF Generation** Continuous assessment Not Presential Duration: 10:00

7			**Unit 3** Duration: 00:30 **Units 3 and 4** Duration: 03:30	
8			**Unit 4** Duration: 00:30 **Unit 4** Duration: 01:30	**RDF generation and linking based on existing ontologies and Knowledge Graphs** Continuous assessment Not Presential Duration: 10:00
9			**Unit 5** Duration: 00:30 **Unit 5** Duration: 03:30	
10			**Unit 5** Duration: 00:30 **Unit 5** Duration: 03:30	**Knowledge-Graph based application development** Continuous assessment Not Presential Duration: 10:00 **Oral presentation** Continuous assessment Not Presential Duration: 02:00
11				
12				
13				
14				
15				
16				
17				**Global exam about course contents** Final examination Not Presential Duration: 02:00

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
3	SPARQL test		No Presential	01:00	10%	/ 10	CB07 CECD01
6	RDF Generation		No Presential	10:00	25%	/ 10	CB07 CECD01
8	RDF generation and linking based on existing ontologies and Knowledge Graphs		No Presential	10:00	15%	/ 10	CECD01 CB07
10	Knowledge-Graph based application development		No Presential	10:00	25%	/ 10	CGI01 CB07 CECD01 CB06
10	Oral presentation		No Presential	02:00	25%	/ 10	CG14 CB06 CB09

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Global exam about course contents		No Presential	02:00	100%	5 / 10	CGI01 CB07 CECD01 CG14 CB06 CB09

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Global exam about the course contents		Face-to-face	02:00	100%	5 / 10	CECD01 CG14 CB06 CB09 CGI01 CB07

7.2. Assessment criteria

The final grade will be based on the evaluation of:

1. The material provided by the student and the interaction on the repository of the course in relation with the work proposed in units 2 and 3. This work will have the weight of 35% (10% for the SPARQL test and 25% for the initial generation of RDF)

2. The material provided by the student and the interaction on the repository of the course in relation with the work proposed in unit 4. This work will have the weight of 15%.

3. The material provided by the student and the interaction on the repository of the course in relation with the work proposed in unit 5. This work will have the weight of 25%.

4. The oral presentation of the complete work. This work will have the weight of 25%.

All this will be evaluated continuously and in the final evaluation activity on the last week of the course. There will be also tests before each lecture so as to ensure that all individual students get enough knowledge about the course topics, even if most of the evaluations are done on groupwork. If the student does not pass this individual exam, he/she will need to do the final exam. There will be also short tests associated to each hands-on delivery so as to ensure as well that the students are knowledgeable about all the work that is being done in groups.

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Ontological Engineering	Bibliography	Ontological Engineering. Gómez-Pérez A, Fernández-López M, Corcho O (2004). Springer-Verlag http://www.springer.com/gp/book/9781852335519
Linked Data applications in Spain	Web resource	http://www.linkeddata.es/

9. Other information

9.1. Other information about the subject

The course will be taught in an intensive manner during the first 10 weeks of the course, so that those students following the continuous evaluation model will have already finished, if successful, by the end of week 10. The course materials will be made available in Moodle, and the online lectures, when needed, will be done using Microsoft Teams. The delivery of all practical work will be done using a GitHub repository.

This course does not contribute explicitly to any SDG, although the application of open data and knowledge graphs to public administrations can be categorised as contributing to SDG16.