



POLITÉCNICA

INTERNATIONAL
CAMPUS OF
EXCELLENCE

COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001



E.T.S. de Ingeniería
Agronómica, Alimentaria y de
Biosistemas

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

203000024 - Bioinformatics Programming Challenges

DEGREE PROGRAMME

20BC - Master Universitario En Biología Computacional

ACADEMIC YEAR & SEMESTER

2020/21 - Semester 1

Index

Learning guide

| | |
|---|----|
| 1. Description..... | 1 |
| 2. Faculty..... | 1 |
| 3. Prior knowledge recommended to take the subject..... | 2 |
| 4. Skills and learning outcomes | 2 |
| 5. Brief description of the subject and syllabus..... | 3 |
| 6. Schedule..... | 5 |
| 7. Activities and assessment criteria..... | 7 |
| 8. Teaching resources..... | 9 |
| 9. Other information..... | 10 |

1. Description

1.1. Subject details

| | |
|---------------------------------------|--|
| Name of the subject | 203000024 - Bioinformatics Programming Challenges |
| No of credits | 3 ECTS |
| Type | Optional |
| Academic year of the programme | First year |
| Semester of tuition | Semester 1 |
| Tuition period | September-January |
| Tuition languages | English |
| Degree programme | 20BC - Master Universitario en Biología Computacional |
| Centre | 20 - E.T.S. De Ingeniería Agronómica, Alimentaria Y De Biosistemas |
| Academic year | 2020-21 |

2. Faculty

2.1. Faculty members with subject teaching role

| Name and surname | Office/Room | Email | Tutoring hours * |
|---|--------------------|-----------------------|-------------------------|
| Mark Denis Wilkinson (Subject coordinator) | B54 CBGP | mark.wilkinson@upm.es | M - 13:15 - 14:00 |

* 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

- Programming competence

4. Skills and learning outcomes *

4.1. Skills to be learned

CE01 - Comprender las bases moleculares y las técnicas experimentales estándares más comunes en las investigaciones ómicas (genómica, transcriptómica, proteómica, metabolómica, interactómica, etc.).

CE04 - Utilizar diferentes bases de datos (incluidos los bigdata), conocer sus estructuras y ontologías, aplicar la estadística a su análisis, siendo capaz de utilizar herramientas de representación y visualización.

CE08 - Capacidad de integrar tecnologías y sistemas propios de la Inteligencia Artificial, con carácter generalista, y en contextos más amplios y multidisciplinares.

CE09 - Capacidad de interpretar los modelos de clasificación supervisada y no supervisada obtenidos al aplicar las técnicas de Aprendizaje Automático para un conjunto de datos.

CG02 - Familiarizarse con el trabajo y los métodos de la Biología Computacional en condiciones reales, adquiriendo la capacidad de diseñar aplicaciones/experimentos de forma independiente y describir, cuantificar, analizar y evaluar críticamente los resultados obtenidos.

CG05 - Que los estudiantes sean capaces de integrar conocimientos en el área de la Biología Computacional, de formular conclusiones, hipótesis o líneas de trabajo a partir de la información disponible, y de formarse una opinión fundamentada sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos.

CT02 - Capacidad para aplicar el método científico para la resolución de problemas de forma efectiva y creativa.

CT03 - Tener compromiso bioético y profesional y respeto por la sostenibilidad ambiental.

4.2. Learning outcomes

RA7 - Adquirir conocimientos para el manejo de técnicas avanzadas de representación de datos y enfoques de integración masiva de datos

RA8 - Adquisición de conocimientos de cuales son las mejores prácticas para la publicación de datos científicos

RA9 - Aprendizaje del uso de interfaces públicas de

* 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

Though biology has quickly become a data-centric science, the practitioners of biology have not understood the importance of their data - in particular, the care that must be taken in annotating it, describing and representing it, and publishing it. Therefore, the greatest challenge for bioinformaticians at this time is related to Finding data, Accessing data, Integrating data, and Reusing data (F.A.I.R.). This course will initially focus on problems related to large-scale data integration. We will study the software tools for capturing data, from files and then directly from the Web. We will learn to manipulate data structures, and how to write clear software that represents complex data structures. We will learn about contemporary approaches to data and knowledge modelling in the life sciences. We will learn how to use these contemporary approaches to do global-scale, cross-resource queries, such that we can integrate the massive amounts of data our biologist colleagues require to interpret their results. Finally, we will learn how to publish scholarly data in a FAIR manner, and how to create Web interfaces to make these massive integrated datasets tractable for biologists.

5.2. Syllabus

1. Data Discovery
2. Data Access
3. Data Integration
4. Data Publishing
5. Data Manipulation & Formatting
6. Formal Knowledge Representation

6. Schedule

6.1. Subject schedule*

| Week | Face-to-face classroom activities | Face-to-face laboratory activities | Distant / On-line | Assessment activities |
|------|-----------------------------------|------------------------------------|--|---|
| 1 | | | GitHub and Proper Code Management Duration: 02:00 | |
| 2 | | | Introduction to the Ruby programming language part 1 Duration: 02:00 | |
| 3 | | | Object Oriented Programming Duration: 02:00 | Assignment 1 - simple objects Continuous assessment Not Presential Duration: 00:00 |
| 4 | | | Automated Web Access Duration: 02:00 | |
| 5 | | | "Representation State Transfer" and RESTful APIs Duration: 02:00 | Assignment #2 - Intensive Data Integration using REST APIs Continuous assessment Not Presential Duration: 00:00 |
| 6 | | | Ruby "gems", Creating documentation, good software publishing and documentation for reusability Duration: 02:00 | |
| 7 | | | BioRuby Level 1 - sequence manipulation Duration: 02:00 | Assignment #3 - locate and visualize insertion sites in a genome Continuous assessment Not Presential Duration: 00:00 |
| 8 | | | BioRuby Level 2 - Other data types, literature, citations, pathways, ontological annotations, etc. Duration: 02:00 | |
| 9 | | | BioRuby Level 2 - Other data types, literature, citations, pathways, ontological annotations, etc. Duration: 02:00 | Assignment #4 Find the orthologues between two genomes Continuous assessment Not Presential Duration: 00:00 |

| | | | | |
|----|--|--|--|--|
| 10 | | | Linked Open Data Duration: 02:00 | |
| 11 | | | SPARQL Query over Global Databases Duration: 02:00 | Assignment #5 Answer biologists questions as SPARQL queries Continuous assessment Not Presential Duration: 00:00 |
| 12 | | | Web Programming and basic Server administration Duration: 02:00 | |
| 13 | | | Contemporary Linked Data solutions Duration: 02:00 | |
| 14 | | | Contemporary Knowledge Representation and FAIR Data Publishing Duration: 02:00 | |
| 15 | | | | Build a Web interface over a linked data database representing Polyadenylation site information Final examination Not Presential Duration: 15:00 |
| 16 | | | | |
| 17 | | | | |

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 | Assignment 1 - simple objects | | No Presential | 00:00 | 20% | 5 / 10 | CG05 CG02 |
| 5 | Assignment #2 - Intensive Data Integration using REST APIs | | No Presential | 00:00 | 20% | 5 / 10 | CT02 CE08 CT03 CE04 |
| 7 | Assignment #3 - locate and visualize insertion sites in a genome | | No Presential | 00:00 | 20% | 5 / 10 | CE09 CG02 CE01 |
| 9 | Assignment #4 Find the orthologues between two genomes | | No Presential | 00:00 | 20% | 5 / 10 | CT02 CE08 CE01 CE04 |
| 11 | Assignment #5 Answer biologists questions as SPARQL queries | | No Presential | 00:00 | 20% | 5 / 10 | CE04 |

7.1.2. Final examination

| Week | Description | Modality | Type | Duration | Weight | Minimum grade | Evaluated skills |
|------|---|----------|---------------|----------|--------|---------------|--|
| 15 | Build a Web interface over a linked data database representing Polyadenylation site information | | No Presential | 15:00 | 100% | 5 / 10 | CG05 CE09 CG02 CT02 CE08 CT03 CE01 CE04 |

7.1.3. Referred (re-sit) examination

| Description | Modality | Type | Duration | Weight | Minimum grade | Evaluated skills |
|---|----------|--------------|----------|--------|---------------|--|
| Build a Web interface over a linked data database representing Polyadenylation site information | | Face-to-face | 15:00 | 100% | 5 / 10 | CG05 CE09 CG02 CT03 CE01 CE04 CT02 CE08 |

7.2. Assessment criteria

For all assignments the criteria for evaluation are:

- 1) Is the answer correct and complete? Biologists will not accept incorrect answers. Regardless of the "elegance" of your software, if the answer is incorrect, the software is not useful.
- 2) Is the output easy to understand (from the perspective of the biologist)?
- 3) Does the software demonstrate that the learning objectives have been achieved? Every assignment has associated competencies - you should demonstrate these competencies by highlighting (in your code documentation) which sections of your code address each of the learning objectives.
- 4) Is the software well-documented and easy to use? Will I be able to understand how it works 2 years later?
- 5) Is the software well-architected?
 - Does it use sensible objects?
 - Does it use sensible and efficient algorithms?
 - Does it consider memory-management (especially for big-data)?
 - Does it re-use, rather than re-invent?

- Is the solution "elegant"?

4) Was it submitted on-time.

8. Teaching resources

8.1. Teaching resources for the subject

| Name | Type | Notes |
|--|--------------|--|
| Principled Web Architecture | Bibliography | http://portal.acm.org/citation.cfm?doid=514183.514185 Roy Fielding's description of his REST design patterns for Web interfaces |
| Data Integration in the era of 'omics | Bibliography | doi: 10.1186/1752-0509-8-S2-I1 General discussion of biological data integration challenges |
| Identification of Orthologues - algorithms | Bibliography | Methods Mol Biol. 2015;1231:203-32. doi: 10.1007/978-1-4939-1720-4_14. |
| Semantic Web Core Technologies | Web resource | https://www.w3.org/RDF/ https://www.w3.org/OWL/ |
| Use of Semantic Web in biomedicine | Bibliography | Samadian S; McManus B; Wilkinson MD. 2012. "Extending and encoding existing biological terminologies and datasets for use in the reasoned semantic web." J. Biomed. Semantics 3:6. |
| SPARQL to query biomedical databases | Bibliography | PLoS Comput Biol. 2016 Jun 23;12(6):e1004989. doi: 10.1371/journal.pcbi.1004989. eCollection 2016. |

| | | |
|---------------------------------|--------------|---|
| The FAIR Principles | Bibliography | https://www.nature.com/articles/sdata201618 |
| FAIR Principles revisited | Bibliography | http://content.iospress.com/articles/information-services-and-use/isu824 |
| The FAIR Principles Implemented | Bibliography | https://peerj.com/articles/cs-110/ |

9. Other information

9.1. Other information about the subject

La asignatura se relaciona con los ODS3, ODS4, y ODS9