



Big Data - knowledge extraction from biological databases

Lecturer: Teresa Nogueira (collaborator at cE3c)

Calendar: July 10-14, 2017

Duration: 36 hours

Schedule: 9h-12h30 and 14h-17h30, Monday-Thursday; 9h-13h and 14h-18h Friday

Overview: During recent decades we have witnessed a great development of bioinformatics that has led to the accumulation of a huge amount of biological information. The Bioinformatics and computational Biology aim at dealing with this large volume of data so that biological information can be extracted, generating scientific knowledge. Handling and mining big data is currently a subject of great interest and importance.

Objectives: This course aim at familiarizing the attendees with the unix environment and shell scripting. The participants will develop and implement querying algorithms in order to generate metadata for analysis.

General Plan:

1. Introduction to Unix/Linux operating systems
2. Unix command line, environment, settings, file system hierarchy, etc.
3. Regular expressions and their use while dealing with biological information
4. Shell commands: metacharacter expansion, redirection, pipes and filters
5. Unix programming: shell scripting, arguments, status, variables, loops, etc.
6. Grep commands
7. Awk scripting language
8. Files transfer protocol
9. Data mining: bibliographic records, gene orthology and synteny of orthologs, biological sequences and annotation data, gene context or correlation, etc.
10. Relational databases and data warehousing
11. Analysis and discussion of case studies

This course can have recognition of 6 ECTS for FCUL PhD students enrolling in it as part of their first doctoral year. For FCUL PhD students only requiring 5 ECTS recognized in their specific PhD programs the last 6 hours of the course are not mandatory and the certificate will be on 'Topics in Big Data – knowledge extraction from biological databases'.

Nº (min, max) students: 10 – 20

Directed to: PhD or MSc students, postdocs, and professionals working in Molecular Biology, Biochemistry, Genetics and related topics.

Minimal formation of students: bachelor degree in Biology, Biochemistry or related areas.

Fee: free for 1st year PhD students in the Doctoral program in Biology (FCUL), Biodiversity, Genetics and Evolution (BIODIV UL; UP) and Biology and Ecology of Global Changes (BEAG UL, UA) when the course counts credits for their formation, in which case the delivery of a final report done after the course is mandatory; 25 € for PhD students from institutions of the PEERS network (cE3c, CFE); 125 € for FCUL Master students and unemployed; 180 € for BTI, BI and other PhD students; 250 € for Professional and postdocs.

When the maximum number of students is reached 10 vacancies will be available for non-paying 1st year PhD students mentioned above, being, by order of preference: 1) cE3c students; 2) BIODIV students (not from cE3c); 3) FCUL students (not from cE3c); 4) BEAG students (not from FCUL).

Deadline for applications: June 9th 2017

To apply send an e-mail to Teresa Nogueira (teresainogueira@gmail.com) with a cv, motivation letter and the following information:

Full Name:

E-mail:

Phone:

Professional activity: Professional/Postdoc, BTI, BI (or other non-post-doc research grant), PhD student (with/ without scholarship), Lic. (Bachelor)/Master student

Academic formation:

PhD student of the 1st year of Doctoral programme BIODIV (FCUL/FCUP), Biologia (FCUL) or BEAG (FCUL or UA)?:

If yes to the above question, PhD student doing the Course to count credits for 1st year?:

PhD student of cE3c or CEF (Centro de Ecologia Funcional):?

If PhD student from another programme/centre, which: