

Titolo: Metodi informatici per la biologia

Titolo (inglese): Informatics methods for Biologists

Obiettivo: Agli studenti sarà offerta una panoramica delle moderne applicazioni informatiche per la biologia: dai diversi linguaggi/ambienti di programmazioni alle differenti basi di dati che i biologi utilizzano durante le loro ricerche. In questo senso, l'obiettivo primario del corso sarà di orientare gli studenti nella scelta delle metodologie più opportune per specifici casi di studio. La prima parte del corso sarà dedicata al consolidamento delle basi conoscitive relative all'informatica e all'utilizzo dei più comuni software per l'organizzazione e editing dei dati (fogli di lavoro e database), la produzione e manipolazione di documenti (editor di testo come Microsoft Word e Latex), alle differenze tra applicazioni e codici proprietari e open-source, alle diverse architetture informatiche ed ai sistemi operative Windows, Unix e Linux.

Inoltre saranno offerti approfondimenti su casi di studio specifici relativi all'ambiente R, al linguaggio C++, al linguaggio MySQL, al linguaggio Python e alla geo-informatica quali prototipi di scelte diverse rispetto a basi dati e finalità proprie dell'analisi statistica, computazionale, della gestione di database, dell'analisi di sequenze genetiche e della manipolazione di dati georeferenziati. Approfondimenti specifici potranno essere impartiti sulla base della richiesta didattica degli studenti del loro livello di formazione.

Aims: An overview of the modern informatics applications in the field of biology will be provided: from the different languages/environments to the different kinds of data that biologists commonly handle through their research activities. Hence, the main aim of the course will be represented by the guidelines in supporting appropriate selection of tools and approaches for specific and different cases of studies. The first part will be dedicated to the consolidation of the basics of computer science and to application of leading software for editing and managing data (spreadsheets, database) and preparation of documents (Microsoft Word, latex). The differences between the architecture of the main operating system (Windows, Unix, Linux) will be investigated.

Working on real case studies, will be presented several tools, such as the R environment, C ++, MySQL language and Geographical Information System (GIS), to be used in relation to the category of data, the purposes of statistical analysis, management data and processing of spatial data.

Specific insights will be given on the basis of students' request taking into account their level of knowledge.

Programma:

1. Introduzione alla struttura e funzionamento del computer.
2. Panoramica delle principali architetture informatiche e dei principali sistemi operativi (Windows, OS, Linux): caratteristiche, vantaggi e svantaggi;
3. Panoramica dei più comuni software per elaborazione testi e dati. Fogli di stile e produzione di documenti e report scientifici;
4. Basi dati: dai documenti di testo ai database alle immagini georeferenziate alle sequenze di codice genetico.
5. Introduzione alla programmazione: cos'è un algoritmo e come lo si costruisce;
6. Pseudocodice e codice: similarità e differenze dei diversi ambienti e linguaggi di programmazione ed analisi;
7. L'ambiente R: caratteristiche, potenzialità, vantaggi e limiti. Approfondimenti su specifici casi di studio;
8. Il linguaggio C++: problematiche, sintassi e utilizzo di base;
9. Il linguaggio MySQL e la gestione dei database: un approccio moderno alla archiviazione, gestione ed utilizzo di grandi quantità di informazioni;
10. Il linguaggio Python e la manipolazione di dati genetici;
11. La geostatica: ArcGIS e Q-Gis. Elaborazione di dati spaziali e manipolazione di immagini da satellite;
12. Approfondimenti specifici da definire durante il corso.

Schedule:

1. Introduction to computer structure and functioning.
2. Overview of the main architectures and major operating systems (Windows, OS, Linux): characteristics, advantages and disadvantages.
3. Overview of the most common software for word processing and data. Stylesheets and production of documents and scientific reports
4. Databases, from text documents to: database, georeferenced images and sequences of genetic code.
5. Introduction to Programming: what is an algorithm and how it is constructed.
6. Pseudo code: similarities and differences of environments and programming languages and analysis.
7. The R environment: characteristics, potential, advantages and limitations. Insights on specific case studies.
8. C ++: problems, syntax and basic use.
9. The language and the MySQL database management: a modern approach to storage, management and use of large amounts of information.
10. Python and the manipulation of genetic data.
11. The geostatic: Q-GIS and ArcGIS. Processing of spatial data handling and satellite images.

12. Insights specific to be defined during the course.

Semestre: secondo

Esame: prova pratica + orale

Recapito telefonico del docente: 0672595854

Recapito email del docente: Tommaso.Russo@Uniroma2.it

Orario e sede di ricevimento: Venerdì dalle 10,30 alle 13,30 presso il Laboratorio di Ecologia Sperimentale ed Acquacoltura