SYLLABUS

Name: *Date science 101 (25-BI-S2-W-DS-AN)*
Name in Polish: **Nauka o danych 101**
Name in English: *Date science 101*

Information on course:

Course offered by department: Faculty of Biological Sciences
Course for department: Faculty of Biological Sciences

Default type of course examination report:

Grading

Language:

English

Short description:

Prerequisites regarding knowledge, skills, and social competences for the course/module:
Basic computer skills, critical and creative thinking.

Student's own work:
- articles reading: 15h
- homework assigements: 20h
- projects preparing: 30h

Description:

Educational aims:
The aim of this course is to provide students with basic concepts, tools and ideas in the data science. Participants will learn all important step in datasets handling, from gathering and processing the data to use it in more advanced analysis.

Course content:
Classes (short introduction in form of lecture + practice):
- introduction to data science: idea, terminology, popular tools, use cases;
- data structures: common file formats, metadata, importance of standards, introduction to Tidy Data, Darwin Core and selected ISO standards;
- data collection: planning data gathering, use of existing databases/datasets, legal and ethical challenges;
- pre-processing: structuring, cleansing, editing, validating; introduction to OpenRefine software and regex;
- connecting and integrating different sources or types of data, cross-checking;
- data mining, documentation of workflows, introduction to Orange software;
- results visualizations and presentation;
- data storage and sharing, and data "afterlife".

Bibliography:

Mandatory and recommended reading list:
Demšar, Janez, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, et al. 'Orange: Data Mining Toolbox in Python'. Journal of Machine Learning Research 14 (2013): 2349–53.
Verborgh, Ruben, and Max De Wilde. Using OpenRefine. Community Experience Distilled. Birmingham Mumbai: Packt Publishing, 2013.
Wickham, Hadley. 'Tidy Data'. Journal of Statistical Software 59, no. 10 (2014). https://doi.org/10.18637/jss.v059.i10.
Wieczorek, John, David Bloom, Robert Guralnick, Stan Blum, Markus Döring, Renato Giovanni, Tim Robertson, and David Vieglais. 'Darwin Core: An Evolving Community-Developed Biodiversity Data Standard'. PLOS ONE 7, no. 1 (6 January 2012): e29715. https://doi.org/10.1371/journal.pone.0029715.

Learning outcomes:

Intended learning outcomes
Student:

K_W08
-knows and can explain basic terms, tools and application cases of data science;

K_U07
plans and properly executes process of creating valuable dataset form data gathering to results presentation;

K_K05
is able to independently look for new tools and techniques necessary to complete given task;

K_K04
is aware of limitations and constraints (ethical, legal and technical) in data usage.

Assessment methods and assessment criteria:

Assessment methods for the intended learning outcomes:
- practical work under supervision,
- homework assignments,
- semester practical assignment.

Credit requirements for individual components of the course/module:
- classes attendance,
- working in groups,
- completed homework assignments,
- semester practical assignment.

Course credits in various terms:

<without a specific program>			
Type of credits	Number	First term	Last term
European Credit Transfer System (ECTS)	3	2023/24-Z	

<without a specific program>

Type of credits	Number	First term	Last term
European Credit Transfer System (ECTS)	3	2023/24-Z	