DATA SCIENCE, MASTER STUDY PROGRAMME

COURSE DESCRIPTIONS

Course name: **STATISTICS**

Number of ECTS credits: 6

Content:

- Probabilistic fundamentals
- Random variables and distributions
- Expectation
- Independence and conditioning
- Approximation of distributions
- Fundamentals of decriptive statistics
- Graphical representations of data
- Population and sample data
- Accuracy of sampling estimators
- Statistical models and their role in data science
- Basics of inferential statistics, parameter estimation
- Basics of hypothesis testing, p-values, goodness of fit measures
- Regression methods, linear regression
- Bayesian paradigm, fundamentals of Bayesian statistic

Course name: **DATABASES FOR BIG DATA**

Number of ECTS credits: 6

Content:

- Relational databases.
- NoSQL databases: Document databases, Graph databases, Column databases, Distributed database patterns.
- Query optimization.
- Concurrency control.
- Crash recovery.

Project tasks will be directly related to the real problems of research projects that are carried out at the university.

Course name: **INTELLIGENT SYSTEMS**

Number of ECTS credits: 6

Content:

Topics covered (including subtopics):

- Fundamental Issues [core*]
- Basic Search Strategies [core]
- Knowledge Based Reasoning [core]
- Advanced Search [elective*]
- Advanced Reasoning [elective]
- Agents [elective]

- Natural Language Processing [elective]
- Machine Learning [elective]
- Planning Systems [elective]
- Robotics [elective]
- Perception [elective]

* - there are 3 core topics that form the basis and 8 elective topics of which 3 will be covered in this course. These 3 elective topics will be selected each year following the current research directions.

Course name: **DATA SCIENCE ETHICS**

Number of ECTS credits: 3

Content:

Modern topics of ethical approach to managing and exploring data. Possible topics include:

- Ethics basic foundation
- Informed consent
- Data ownership
- Privacy
- Anonymity
- Data validity
- Algorithmic fairness
- Societal consequences
- Code of ethics

Course name: **SELECTED TOPICS IN DISCRETE MATHEMATICS**

Number of ECTS credits: 6

Content:

The most current research topics in discrete mathematics will be taught, which may include, among others, the following topical subsections:

- Theory of configurations,
- Graph theory,
- Algebraic methods in graph theory,
- Large network theory and analysis,
- Learning on Networks,
- Random graph walks,
- World-Wide Web as a graph.

At seminars concrete examples of the problems that are encounter in the implementation P1-0285 research program will be addressed (eg data analysis under EEG measurements).

Course name: **SELECTED TOPICS IN NUMERICAL MATHEMATICS**

Number of ECTS credits: 6

Content:

Basic actual research topics are considered from several fields of numerical mathematics, such as:

- Approximation of functions.
- Numerical analysis of ordinary differential equations

- Numerical analysis of partial differential equations
- Numerical methods for large linear systems and
- Numerical methods for large eigenvalue problems.
- Large scale numerical optimization
- Bezier curves and surfaces

Course name: **SELECTED TOPICS IN INFORMATION VISUALISATION**
Number of ECTS credits: 6

Content:
Classical and modern research topics in the field of information visualisation. Possible topics include:

- appropriate allocation of visual attributes to data variables,
- designing with color and luminance contrast,
- the psychology of human vision and perception,
- visual analytics,
- interaction,
- storytelling,
- text visual analytics,
- big data visualization,
- uncertainty visualization,
- network visualization,
- cartographic visualization,
- animation and time series visualization

Course name: **DATA PRACTICUM I**
Number of ECTS credits: 3

Content:

- Introduction to basic syntax and programming in R.
- Data structures and data manipulation.
- Base functions.
- Efficient and reproducible data managmenent and statical analysis in R.
- Graphical representation using base functions and ggplot2.
- Development and deployment of interactive reports and of web applications using R, Rmarkdown and the shiny package
- R code development and optimization: Common errors that make the code inefficient; Annotation and versioning control; Testing; Debugging; Profiling and performance measurement; Memory management; Computing for computationally intensive methods (integration of C++ code in R functions using the Rcpp package, parallel computing);
- R Package development: Package structure; Generic functions and methods; Checking and building packages; Package documentation; Version control with GitHub
- Data analysis applications in R: Summary statistics; Parametric and non-parametric significance tests; Regression & GLMs (linear, logistic, multiple); Mixed models

At exercises students will use real big data from research projects carried out under the research programs P1-0285, P1-0404 and P1-0294.

Course name: **DATA ENGINEERING AND DISTRIBUTED INFORMATION SYSTEMS**

Number of ECTS credits: 6

Content:

Classical and modern research topics in the field of data inginering and distributed systems. Possible topics include:

- advanced tools for data manipulation, data organization and storge;
- software product development technologies such as: computer aided software engineering (CASE), collaboration tools for software development (e.g. versioning systems, ticketing systems);
- data communication / computer networks;
- client / server architecture, peer-to-peer systems;
- web applications and web services;
- modern applications of distributed information systems.

Course name: **GRAPH ALGORITHMS**

Number of ECTS credits: 6

Content:

Classical and modern research topics in the field of algorithmic graph theory. Possible topics include:

- Basic graph definitions, representations of graphs
- Trees and forests, their basic properties
- Breadth-first search and depth-first search
- Eulerian tours and Hamiltonian cycles
- Minimum spanning trees, shortest paths
- Bipartite graphs, graph colorings
- Matchings in bipartite graphs
- Network flows
- Planar graphs
- Approximation algorithms (e.g., for the traveling salesman problem and for the vertex cover problem)
- Chordal graphs, interval graphs, perfect graphs
- Graphs of bounded tree-width
- Graphs of bounded clique-width