KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

Projekt współfinansowany przez
Unię Europejską w ramach
Europejskiego Funduszu
Społecznego

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Course title	ECTS code
Data bases & big data	13.3.1302

Name of unit administrating study

null

Studies

faculty	field of study		
Wydział Chemii	Chemia	type	drugiego stopnia
		form	stacjonarne
		specialty	Digital Chemistry
		specialization	wszystkie

Teaching staff

prof. dr hab. Tomasz Puzyn; mgr Maciej Gromelski; dr Alicja Mikołajczyk

Forms of classes, the realization and number of hours	ECTS credits
Forms of classes	2
Laboratory classes	Laboratory classes - 30 h
The realization of activities	(tutorial classes – 10 h,
classroom instruction	student's own work – 10 h)
Number of hours	
Laboratory classes: 30 hours	Total: 50 h - 2 ECTS

The academic cycle

2023/2024 summer semester

Type of course	Language of instruction
an elective course	English

Teaching methods	Form and method of assessment and basic criteria for eveluation or examination requirements
Practical laboratory work – students will learn aspects of Big Data and solve different exercises faced by the teacher (on computer stations).	**Final evaluation**
	Graded credit
Introduction with the use of the multimedia presentation on the basic issues of data science and Big Data.	**Assessment methods**
	Laboratory classes – short tests and reports
	The basic criteria for evaluation

C. The basic criteria for evaluation or exam requirements - the correctness of the reports on assigned projects, the final grade of the lab. is based on the partial grades received from each report and presentation of the final project; failure to complete the experimental part means failing the laboratory exercises Laboratory classes: positive note from all short tests and reports, final note is an average from notes from all tests 91-100%: 5.0 81-90%: 4.5 71-80%: 4.0 61-70%: 3.5 51-60%: 3.0 < 51%: 2.0 a. positive grade for the written test consisting of 15-20 open questions covering only the issues mentioned in the lecture issues and the issues discussed during the auditorium exercises; the exam may be taken by a student who has passed the auditorium and laboratory exercises. Assessment criteria in accordance with the University of Gdańsk Study Regulations Lab classes: the arithmetic mean of partial grades received during the semester for written reports on laboratory exercises and presentation of the final assignment; the main criteria for evaluation of reports are the correct answers to the questions in the exercise instructions. Lectures: passing the final exam in the form of a multiple-choice question test (a score of 50% or more required to pass the test).

Method of verifying required learning outcomes

Required courses and introductory requirements

A. Formal requirements
lack

B. Prerequisites
lack

Aims of education

familiarize students with the main aspects of databases & Big Data
familiarize students with Python programming language and Apache Spark to analyze Big Data
familiarize students with SparkSQL, DataFrames and DataSets
familiarize students with machine learning techniques using SparkML

Course contents

concept of large databases and BigData, basics of big datasets engineering, big data hardware infrastructure (local and cloud), MapReduce algorithm, introduction to Python programming language, data analysis in Python, machine learning (supervised and unsupervised methods), introduction to Apache Spark and Hadoop
setting up working environment (Python, Spark) and Big Data datasets, Spark basics, Resilient distributed datasets, RDDs examples and exercises, introduction to SparkSQL, SQL commands exercises, Spark MLlib (linear regression and decision trees with Spark ML)

Bibliography of literature

Literature required to pass the course
Apache Spark and PySpark documentation - https://spark.apache.org/docs/
Python documentation - https://docs.python.org/3/
 Extracurricular readings
 M. Bowles - Machine Learning with Spark™ and Python®: Essential Techniques for Predictive Analytics

The learning outcomes (for the field of study and specialization)	Knowledge
K_W02 depth knowledge in the field of basic chemistry K_W08 demonstrates in-depth knowledge of theoretical computational and IT methods used to solve problems in	Students: know main concepts of Big Data; understand the structure and properties of databases; understand the hardware requirements and differences in the infrastructure for big data; understand the MapReduce algorithm and its mapper and reduction functions; understand basics of the Python programming language (types, data structures, functions, libraries), know essential methods and libraries used in data analysis and machine learning in Python; know Apache Spark and Hadoop

chemistry K_U01 plans and implements chemical experiments of extended complexity K_U08 prepares and presents oral presentations in various fields of chemistry in Polish and English, using acquired knowledge and skills as well as basic sources of scientific information K_U09 has deepened ability to prepare various forms of oral presentations on chemistry in Polish and English K_K06 undertakes research tasks consciously and responsibly, understanding the social aspects of the practical application of the acquired knowledge and skills and the responsibility related to it	engines and its modules
	Skills
	Students: present plainly – in both speech and writing – correct argumentation related to data science problems; write and explain basic source code in Python programming language to resolve given data-related problems; use Spark Resilient Distributed Datasets to process and analyze large data sets; understand and write proper SQL syntax statements to process large datasets; use Spark MLLib to perform machine learning tasks; interpret information, formulates conclusions and explain opinions.
	Social competence
	Students: understand need for learning, inspire other for learning; cooperate in group, taking different roles; exhibit creativity in determination of priorities necessary for realization of different tasks; understand social aspects of practical use of knowledge and abilities as well as connected with them responsibility.

Contact

tomasz.puzyn@ug.edu.pl