Big data technologies - course description

General information

Course name	Big data technologies
Course ID	11.3-WE-BizElP-TechBigData-Er
Faculty	Faculty of Computer Science, Electrical Engineering and Automatics
Field of study	E-business
Education profile	practical
Level of studies	First-cycle Erasmus programme
Beginning semester	winter term 2021/2022

Course information

Semester	3
ECTS credits to win	5
Course type	obligatory
Teaching language	english
Author of syllabus	• dr hab. inż. Artur Gramacki, prof. UZ

Classes forms

The class form	Hours per semester (full-time)	Hours per week (full-time)	Hours per semester (part-time)	Hours per week (part-time)	Form of assignment
Lecture	30	2	-	-	Credit with grade
Project	30	2	-	-	Credit with grade

Aim of the course

Teaching students how to choose the right data analysis techniques depending on the scale of the problem being considered and the type of analysis being carried out.

Teaching students to work using modern platforms for data storage and processing.

Teaching students selected techniques to analyze large data sets, mainly textual.

Prerequisites

Introduction to databases, Basics of statistics

Scope

Big Data: An introduction to processing large amounts of data.

Non-relational databases: Reminder of the basic issues related to relational databases. Advantages and disadvantages of these databases. Basic problems related to the use of relational databases to store and process larger and larger amounts of increasingly distributed data. Horizontal and vertical scaling of databases. A new concept of databases not based on the traditional relational model. CAP and BASE theory. Aggregate data models. Key-value, column, document and graph databases. Database replication. Sharing resources in databases. Map-Reduce methodology. Presentation of a few selected non-relational database systems (e.g. MongoDB, Cassandra, Redis, Neo4J, Oracle NoSQL Database).

Selected IT systems: Large-scale business analytics: modern solutions used for transmission, storage and processing of large data sets. Basics of data processing using convolutional neural networks (CNN). Tensorflow and Keras libraries. Working in the Google Colaboratory cloud environment.

Teaching methods

lecture: conventional lecture

project: work in groups, practical classes

Learning outcomes and methods of theirs verification

Outcome description	Outcome symbols	Methods of verification	The class form
Knows the databases known as NoSQL, is able to define their basic features and types, is able to use the most popular systems of such databases.		• an examination test with score scale • an ongoing monitoring during classes	• Lecture
Is able to define the concept of Text Mining and give typical examples of tasks in this area of knowledge.		• an examination test with score scale	• Lecture
Knows the techniques of searching for text information and creating their ranking.		• an examination test with score scale	• Lecture

Outcome description	Outcome symbols	Methods of verification	The class form
Is able to perform selected statistical analyzes of text documents.		• an ongoing monitoring during classes	• Project
Is able to characterize the features of modern Big Data platforms.		• an examination test with score scale	• Lecture
Is able to use the tools offered by the Elasticsearch platform to perform advanced data analysis and real-time exploration.		• an ongoing monitoring during classes	• Project

Assignment conditions

Recommended reading

1. Pramod J. Sadalage and Martin Fowler: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence,2012
2. Dan Sullivan: NoSQL for Mere Mortals,2015
3. Francois Chollet: Deep Learning with Python, Helion, 2017
4. Tensorflow and Keras docs: https://www.tensorflow.org/guide

Further reading

Notes

Modified by dr hab. inż. Artur Gramacki, prof. UZ (last modification: 14-07-2021 13:03)

Generated automatically from SylabUZ computer system