

Big data and business intelligence - course description

General information	
Course name	Big data and business intelligence
Course ID	11.3-WE-INFD-BDial-Er
Faculty	Faculty of Computer Science, Electrical Engineering and Automatics
Field of study	Computer Science
Education profile	academic
Level of studies	Second-cycle Erasmus programme
Beginning semester	winter term 2021/2022

Course information	
Semester	2
ECTS credits to win	5
Course type	obligatory
Teaching language	english
Author of syllabus	

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
Laboratory	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.

Elements of Text Mining: Introduction to Text Mining. Pre-processing of text documents. Stemming algorithms. Keyword searching. Organization of documents in the form of a term-document matrix (TDM). Selected elements of linear algebra and their application to Text Mining. Grouping and classifying of text documents. Create document summaries. Wordclouds. Sentiment analysis. Selected IT systems and libraries for Text Mining.

Teaching methods

Lecture, laboratory exercises.

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.		<ul style="list-style-type: none">• a multiple choice and open questions test• an oral response	<ul style="list-style-type: none">• Lecture
Is able to define the concept of Text Mining and give typical examples of tasks in this area of knowledge.		<ul style="list-style-type: none">• a multiple choice and open questions test• an oral response	<ul style="list-style-type: none">• Lecture
Can make simple analysis using convolutional neural networks (CNN).		<ul style="list-style-type: none">• a preparation of a project	<ul style="list-style-type: none">• Laboratory
Is able to use the most popular NoSQL systems.		<ul style="list-style-type: none">• an ongoing monitoring during classes	<ul style="list-style-type: none">• Laboratory

Outcome description	Outcome symbols	Methods of verification	The class form
Can characterize convolutional neural networks and compare them to the classical neural networks.		<ul style="list-style-type: none"> • a multiple choice and open questions test • an oral response 	<ul style="list-style-type: none"> • Lecture
Is able to make simple text analysis using a selected IT solutions.		<ul style="list-style-type: none"> • a preparation of a project 	<ul style="list-style-type: none"> • Laboratory

Assignment conditions

Lecture – the passing condition is to obtain a positive mark from the final test.

Laboratory – the passing condition is to obtain positive marks from all laboratory exercises to be planned during the semester.

Calculation of the final grade: lecture 50% + laboratory 50%

Recommended reading

1. Daniel Larose: Discovering Knowledge in Data: An Introduction to Data Mining, Wiley, 2014
2. Zdravko Markov, Daniel Larose: Data Mining the Web. Patterns in Web Content, Structure and Usage, Wiley, 2007
3. Francois Chollet: Deep Learning. Deep Learning with Python, Manning Publications Co., 2018
4. Machale W. Berry, Murray Browne: Understanding Search Engines. Mathematical Modeling and Text Retrieval, SIAM, 1999
5. Lars Elden: Matrix Methods in Data Mining and Pattern Recognition, SIAM, 2007
6. Python, R, Keras and TensorFlow documentation

Further reading

Notes

Modified by dr hab. inż. Artur Gramacki, prof. UZ (last modification: 08-09-2021 19:00)

Generated automatically from SyllabUZ computer system