

Technologie big data - opis przedmiotu

Informacje ogólne	
Nazwa przedmiotu	Technologie big data
Kod przedmiotu	11.3-WE-BEP-TBD
Wydział	Wydział Informatyki, Elektrotechniki i Automatyki
Kierunek	Biznes elektroniczny
Profil	praktyczny
Rodzaj studiów	pierwszego stopnia z tyt. inżyniera
Semestr rozpoczęcia	semestr zimowy 2021/2022

Informacje o przedmiocie	
Semestr	3
Liczba punktów ECTS do zdobycia	5
Typ przedmiotu	obowiązkowy
Język nauczania	polski
Sylabus opracował	

Formy zajęć					
Forma zajęć	Liczba godzin w semestrze (stacjonarne)	Liczba godzin w tygodniu (stacjonarne)	Liczba godzin w semestrze (niestacjonarne)	Liczba godzin w tygodniu (niestacjonarne)	Forma zaliczenia
Wykład	30	2	18	1,2	Zaliczenie na ocenę
Projekt	30	2	18	1,2	Zaliczenie na ocenę

Cel przedmiotu

Nauczenie studentów doboru odpowiednich technik analizy danych w zależności od skali rozpatrywanego problemu oraz rodzaju przeprowadzanej analizy.

Nauczenie studentów pracy z wykorzystaniem nowoczesnych platform do składowania i przetwarzania danych.

Zapoznanie studentów z wybranymi technikami analizowania dużych zbiorów danych, głównie tekstowych.

Wymagania wstępne

Bazy danych.

Znajomość podstaw statystyki.

Zakres tematyczny

Big Data: wprowadzenie do zagadnienia przetwarzania wielkich ilości danych.

Nierelacyjne bazy danych: Przypomnienie podstawowych zagadnień związanych z relacyjnymi bazami danych. Zalety i wady tych baz danych. Podstawowe problemy związane z wykorzystaniem relacyjnych baz danych do składowania i przetwarzania coraz większych ilości danych coraz bardziej rozproszonych. Skalowanie poziome oraz pionowe baz danych. Nowa koncepcja baz nie opartych o tradycyjny model relacyjny. Teoria CAP oraz BASE. Agregacyjne modele danych. Bazy danych typu klucz-wartość, kolumnowe, dokumentowe, grafowe. Replikacja baz danych. Współdzielenie zasobów w bazach danych. Metodologia Map-Reduce. Przedstawienie kilku wybranych systemów baz danych nierelacyjnych (np. MongoDB, Cassandra, Redis, Neo4J, Oracle NoSQL Database).

Wybrane systemy informatyczne: Analityka biznesowa na dużą skalę: nowoczesne rozwiązania wykorzystywane do przesyłania, składowania oraz przetwarzania dużych zbiorów danych. Architektura nowoczesnych systemów do składowania i przetwarzania Big Data na przykładzie platformy Elasticsearch. Analityka danych tekstowych w czasie rzeczywistym z wykorzystaniem platformy ElasticSearch. Podstawy przetwarzania danych z wykorzystaniem sieci splotowych (CNN, Convolutional Neural Networks). Biblioteka Keras oraz Tensorflow. Praca w środowisku chmurowym Google Colaboratory.

Metody kształcenia

Wykład konwencjonalny, zadanie projektowe.

Efekty uczenia się i metody weryfikacji osiągnięcia efektów uczenia się

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Student zna różnice pomiędzy tradycyjnymi bazami relacyjnymi a bazami określanymi mianem baz NoSQL. Potrafi wybrać najwłaściwsze rozwiązanie do realizacji określonego zadania.	K_W13	- przygotowanie projektu - test końcowy	- Wykład - Projekt
Student zna bazy danych określane mianem NoSQL, potrafi zdefiniować ich podstawowe cechy i rodzaje, potrafi używać najpopularniejsze systemy takich baz danych	K_W01	- przygotowanie projektu - test końcowy	- Wykład - Projekt
Student potrafi scharakteryzować cechy nowoczesnych platform Big Data.	K_U16	test końcowy	Wykład

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Student potrafi w praktyczny sposób wykorzystać narzędzia oferowane przez platformę Elasticsearch w celu przeprowadzenia zaawansowanej analizy danych oraz ich eksploracji w czasie rzeczywistym.	K_W01	przygotowanie projektu	Projekt

Warunki zaliczenia

Wykład - warunkiem zaliczenia jest uzyskanie pozytywnej oceny z testu końcowego

Projekt - warunkiem zaliczenia jest uzyskanie oceny pozytywnej w wykonanego projektu

Składowe oceny końcowej = wykład: 50% + projekt: 50%

Literatura podstawowa

1. Larose D.T.: Metody i modele eksploracji danych, PWN, Warszawa, 2008
2. Markov Z., Larose D.T.: Eksploracja zasobów internetowych, PWN, Warszawa, 2009
3. Sadalage P. J., Fowler M.: NoSQL. Kompendium wiedzy, 2014
4. Gormley C., Tong Z.: Elasticsearch: The Definitive Guide, 2015
5. Francois Chollet: Deep Learning. Praca z językiem Python i biblioteką Keras, Helion, 2019
6. Dokumentacja systemu R

Literatura uzupełniająca

Uwagi

Zmodyfikowane przez dr hab. inż. Artur Gramacki, prof. UZ (ostatnia modyfikacja: 03-05-2021 22:34)

Wygenerowano automatycznie z systemu SylabUZ