

SYLABUS ZAJ /GRUPY ZAJ

Dane ogólne:

Jednostka organizacyjna:	Katedra Informatyki				
Kierunek studiów:	Informatyka				
Specjalno /Specjalizacja:	Inżynieria oprogramowania				
Nazwa zaj / grupy zaj :	Big data i hurtownie danych				
Course / group of courses:	Big Data and Data Warehouses				
Forma studiów:	stacjonarne				
Nazwa katalogu:	WP-IN-I-20/21Z-IO				
Nazwa bloku zaj :					
Kod zaj /grupy zaj :	105907	Kod Erasmus:			
Punkty ECTS:	4	Rodzaj zaj :		obowiązkowy	
Rok studiów:	3	Semestr:		5	
Rok	Semestr	Forma zaj	Liczba godzin	Forma zaliczenia	ECTS
3	5	LO	30	Zaliczenie z ocen	2
		W	15	Egzamin	2
Razem			45		4
Koordinator:	magister inżynier Tomasz Potempa				
Prowadzący zajęcia:					
Język wykładowy:	semestr: 5 - język polski				

Objaśnienia:

Rodzaj zaj : obowiązkowe, do wyboru.

Forma prowadzenia zaj : W - wykład, - wiczenia audytoryjne, L - lektorat, S – seminarium/ zajęcia seminaryjne, P - wiczenia praktyczne (w tym zajęcia wf), M - wiczenia specjalistyczne (medyczne/ kliniczne), LO – wiczenia laboratoryjne, LI - laboratorium informatyczne, ZTI - zajęcia z technologii informacyjnych, P – wiczenia projektowe, ZT – zajęcia terenowe, T - wiczenia terenowe na obozach programowych, SK - samokształcenie (i inne), PR - praktyka zawodowa

Dane merytoryczne

Wymagania wstępne:			
1. Podstawowa znajomość matematyki dyskretnej w zakresie algebry zbiorów oraz algebry relacji. 2. Znajomość podstaw relacyjnych baz danych. 3. Znajomość podstawowych zagadnień statystyki. 4. Znajomość systemów operacyjnych i podstaw użytkowania komputerów. 5. Znajomość języka angielskiego w stopniu umożliwiającym studiowanie literatury fachowej.			
Szczegółowe efekty uczenia się			
Lp.	Student, który zaliczył zajęcia zna i rozumie/potrafi/jest gotowy do:	Kod efektu dla kierunku studiów	Sposób weryfikacji efektu uczenia się
1	Zna i rozumie podstawową terminologię, pojęcia oraz metody w zakresie analizy danych.	IN1_W01	kolokwium
2	Zna cele i zasady pracy systemów analitycznych oraz systemów przetwarzania dużych zbiorów danych.	IN1_W07, IN1_W08, IN1_W04	kolokwium
3	Zna wybrane, przykładowe narzędzia przetwarzania i analizy danych.	IN1_W07, IN1_W08, IN1_W04	kolokwium

4	Umie przygotować dane oraz przeprowadzić na nich proste oraz bardziej złożone analizy.	IN1_U05, IN1_U12	wykonanie zadania
5	Umie posługiwać się i wykorzystywać do analizy danych przykładowe środowiska i systemy.	IN1_U05, IN1_U12	wykonanie zadania
Stosowane metody osiągnięcia zakładanych efektów uczenia się (metody dydaktyczne)			
metody praktyczne (Pomocniczym form zajęć jest laboratorium komputerowe.), metody podające (Wykład problemowy, metoda (analiza) przypadków, pokaz, prezentacja.)			
Kryteria oceny i weryfikacji efektów uczenia się			
wiedza: ocena kolokwium (Kolokwium)			
umiejętności: ocena wykonania zadania (Wykonanie zadania)			
Warunki zaliczenia			
Zaliczenie ćwiczeń laboratoryjnych oraz zdanie kolokwium. Oceny wystawiane zgodnie z aktualnym regulaminem studiów w PWSZ w Tarnowie.			
Treści programowe (opis skrócony)			
1. Podstawowe koncepcje hurtowni danych oraz systemów przetwarzania dużych zbiorów danych. Przetwarzanie operacyjne i analityczne. 2. Projektowanie hurtowni danych. 3. Architektury gromadzenia dużych zbiorów danych. Modele danych typu Big Data. Metody pozyskiwania i przetwarzania dużych zbiorów danych. 4. Zaawansowane elementy języka SQL w przetwarzaniu danych. 5. Klasyczne metody statystyczne w analizie danych. 6. Metody sztucznej inteligencji w przetwarzaniu danych. Eksploracja danych. Metody reprezentacji i przetwarzania wiedzy.			
Content of the study programme (short version)			
1. Fundamental concepts of data warehouse and big data processing systems. Operational and analytical processing. 2. Designing a data warehouses. 3. Architectures of collecting big data sets. Big Data data models. Methods of obtaining and processing large data sets. 4. Advanced elements of the SQL language in data processing. 5. Classical statistical methods in data analysis. 6. Artificial intelligence methods in data processing. Data mining. Methods of representation and processing of knowledge.			
Treści programowe			
			Liczba godzin
Semestr: 5			
Forma zajęć : wykład			
1. Koncepcja hurtowni danych oraz systemów przetwarzania dużych zbiorów danych. Przetwarzanie operacyjne i analityczne. Architektura hurtowni danych. Relacyjny model danych. Normalizacja. Model wielowymiarowy. Denormalizacja. Podstawowe operacje języka SQL. Operacje na wielowymiarowych strukturach danych. 2. Klasyfikacja wymiarów w hurtowni danych. Projektowanie tabel faktów. Miary. Projektowanie koncepcyjne. Model logiczny. Schemat o strukturze gwiazdy, płatka śniegu, konstelacji. Model fizyczny. 3. Architektury gromadzenia dużych zbiorów danych. Modele danych typu Big Data. Model In-Memory oraz Computing. Metody pozyskiwania i przetwarzania dużych zbiorów danych. 4. Zaawansowane elementy języka SQL. Definiowalne funkcje agregujące. Partycjonowanie. Kursory. 5. Podstawowe pojęcia statystyki opisowej. Analiza szeregów rozdzielczych w R. Podstawy analizy szeregów czasowych. Wybrane metody prognozowania. Korelacja i regresja. Regresja liniowa. 6. Etapy eksploracji danych. Metody grupowania danych. Miary jakości grupowania danych. Klasyfikacja. Etapy konstruowania klasyfikatora. Kryteria oceny algorytmów klasyfikacyjnych. Budowanie drzew decyzyjnych w R. Zależności asocjacyjne.			15
Forma zajęć : wiczenia laboratoryjne			
1. Koncepcja hurtowni danych oraz systemów przetwarzania dużych zbiorów danych. Przetwarzanie operacyjne i analityczne. Architektura hurtowni danych. Relacyjny model danych. Normalizacja. Model wielowymiarowy. Denormalizacja. Podstawowe operacje języka SQL. Operacje na wielowymiarowych strukturach danych. 2. Klasyfikacja wymiarów w hurtowni danych. Projektowanie tabel faktów. Miary. Projektowanie koncepcyjne. Model logiczny. Schemat o strukturze gwiazdy, płatka śniegu, konstelacji. Model fizyczny. 3. Architektury gromadzenia dużych zbiorów danych. Modele danych typu Big Data. Model In-Memory oraz			30

<p>Computing. Metody pozyskiwania i przetwarzania dużych zbiorów danych.</p> <p>4. Zaawansowane elementy języka SQL. Definiowalne funkcje agregujące. Partycjonowanie. Kursory.</p> <p>5. Podstawowe pojęcia statystyki opisowej. Analiza szeregów rozdzielczych w R. Podstawy analizy szeregów czasowych. Wybrane metody prognozowania. Korelacja i regresja. Regresja liniowa.</p> <p>6. Etapy eksploracji danych. Metody grupowania danych. Miary jakości grupowania danych. Klasyfikacja. Etapy konstruowania klasyfikatora. Kryteria oceny algorytmów klasyfikacyjnych. Budowanie drzew decyzyjnych w R. Zależności asocjacyjne.</p>	30
Literatura	
Podstawowa	
1. Jeffrey D. Ullman, Jennifer Widom, Podstawowy kurs systemów baz danych, Helion, Gliwice 2011	
Agnieszka Chodkowska-Gyurics, Hurtownie danych. Teoria i praktyka., Naukowe PWN SA, Warszawa 2014	
Chris Todman, Projektowanie hurtowni danych,, Helion, Gliwice 2011	
Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom, Systemy baz danych. Pełny wykład., WNT, Warszawa, 2006	
Joe Celko, SQL zaawansowane techniki programowania, Naukowe PWN, Warszawa 2008	
Karol Kukuła, Elementy statystyki w zadaniach, Naukowe PWN, Warszawa 1998	
Kuis Torgo, Data Mining with R. Learning with Case Studies, CRC Press 2010	
Nathan Marz, James Warren., Big Data. Najlepsze praktyki budowy skalowalnych systemów obsługi danych w czasie rzeczywistym., Helion 2016	
Przemysław Biecek, Przewodnik po pakiecie R, wyd. 3, Oficyna Wydawnicza GiS, wyd. 3, Wrocław 2014	
Russell Jurney, Zwinna analiza danych. Apache Hadoop dla każdego, Helion 2015	
Stanisław Osowski, Metody i narzędzia eksploracji danych, BTC, Legionowo 2013	
Tadeusz Morzy, Ekploracja danych. Metody i algorytmy., PWN, Warszawa 2013	
Tom White, Hadoop., Hadoop. Kompletne przewodnik. Analiza i przechowywanie danych., Helion 2015	
Tomasz Górecki, Podstawy statystyki z przykładami w R, BTC, Legionowo 2011	
Vidette Poe, Patricia Klauer, Stephen Brobst., Tworzenie hurtowni danych, Naukowo Techniczne, Warszawa 2000	
Wacława Starzyńska, Statystyka praktyczna, Naukowe PWN, Warszawa 2000	
Uzupełniająca	

Dane jako ciowe

Przyporządkowanie zajęć/grup zajęć do dyscypliny naukowej/artystycznej	informatyka techniczna i telekomunikacja
Sposób określania liczby punktów ECTS	
Forma nakładu pracy studenta (udział w zajęciach, aktywność, przygotowanie sprawozdania, itp.)	Obciążenie studenta [w godz.]
Udział w zajęciach	45
Konsultacje z prowadzącym	3
Udział w egzaminie	2
Bezpośredni kontakt z nauczycielem - inne	0
Przygotowanie do laboratorium, ćwiczenia, zajęcia	20
Przygotowanie do kolokwium i egzaminu	15
Indywidualna praca własna studenta z literatury, wykładami itp.	15
Inne	0

Sumaryczne obciążenie prac studenta	100	
Liczba punktów ECTS		
Liczba punktów ECTS	4	
Zajęcia wymagające bezpośredniego udziału nauczyciela akademickiego	L. godzin	ECTS
	50	2,0
Zajęcia o charakterze praktycznym	L. godzin	ECTS
	77	3,1

1 godz = 45 minut; 1 punkt ECTS = 25-30 godzin

W sekcji 'Liczba punktów ECTS' suma punktów ECTS zaj wymagaj cych bezpo redniego udziału nauczyciela akademickiego i o charakterze praktycznym mo e si ró ni od ł cznej liczby punktów ECTS dla zaj /grup zaj .