

Karta przedmiotu

Nazwa i kod przedmiotu	Data bases & big data (Ćw. laboratoryjne), PG_00117811						
Kierunek studiów	Chemia (O)						
Data rozpoczęcia studiów	październik 2024 r.	Rok akademicki realizacji przedmiotu			2024/2025		
Poziom kształcenia	II stopnia	Grupa zajęć			Grupa zajęć fakultatywnych		
Forma studiów	stacjonarne	Sposób realizacji			na uczelni		
Rok studiów	1	Język wykładowy			angielski		
Semestr studiów	2	Liczba punktów ECTS			2.0		
Profil kształcenia	ogólnoakademicki	Forma zaliczenia			zaliczenie		
Jednostka prowadząca	Wydział Chemii -> Katedra Chemii i Radiochemii Środowiska -> Pracownia Chemoinformatyki Środowiska						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		mgr inż. Michał Kałapus				
	Prowadzący zajęcia z przedmiotu						
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	0.0	0.0	30.0	0.0	0.0	30
	W tym liczba godzin zajęć na odległość: 0.0						
Aktywność studenta i liczba godzin pracy	Aktywność studenta	Udział w zajęciach dydaktycznych, objętych planem studiów		Udział w konsultacjach		Praca własna studenta	RAZEM
	Liczba godzin pracy studenta	30		5.0		15.0	50
Cel przedmiotu	Celem przedmiotu jest zapoznanie studentów z podstawami przetwarzania dużych zbiorów danych (Big Data) z wykorzystaniem technologii MapReduce i PySpark. Studenci nauczą się projektować, implementować i optymalizować algorytmy przetwarzania danych w tych technologiach, a także korzystać z zasobów chmurowych (AWS, Google Cloud) do ich uruchamiania.						

Efekty uczenia się przedmiotu	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[CHEMMU2_K06] W sposób świadomy i odpowiedzialny podejmuje się realizację zadań badawczych, rozumiejąc społeczne aspekty praktycznego zastosowania zdobytej wiedzy i umiejętności oraz związaną z tym odpowiedzialność.	Student potrafi zidentyfikować i przeanalizować społeczne, etyczne oraz prawne aspekty związane z wykorzystaniem dużych zbiorów danych, włączając w to dane chemiczne, w szczególności w kontekście ochrony prywatności, bezpieczeństwa danych oraz potencjalnych skutków społecznych wynikających z analizy danych. Student potrafi uwzględnić te aspekty w planowaniu i realizacji projektów badawczych z wykorzystaniem metod obliczeniowych i informatycznych.	[SK5] realizacja zadania problemowego [SK8] obserwacja samodzielnej lub zespołowej pracy studenta
	[CHEMMU2_U01] Planuje i realizuje eksperymenty chemiczne o pogłębionym stopniu złożoności.	Student potrafi zaprojektować i przeprowadzić analizę dużych zbiorów danych pochodzących z eksperymentów, w tym chemicznych, o pogłębionym stopniu złożoności. Student potrafi wykorzystać metody obliczeniowe i informatyczne do przetwarzania, analizy i interpretacji danych (MapReduce oraz PySpark), uwzględniając przy tym specyfikę eksperymentu oraz potencjalne źródła błędów.	[SU1] wypowiedź ustna/rozmowa/diskusja
	[CHEMMU2_W02] Operuje pogłębioną wiedzą w zakresie podstawowych działów chemii.	Student potrafi zastosować pogłębioną wiedzę z zakresu podstawowych działów chemii (chemia organiczna, nieorganiczna, fizyczna, analityczna) do analizy dużych zbiorów danych chemicznych. Student potrafi zidentyfikować związki chemiczne, przewidzieć ich właściwości oraz reaktywność, a także ocenić wpływ różnych czynników na przebieg reakcji chemicznych.	[SW5] realizacja zadania problemowego
	[CHEMMU2_W08] Wykazuje się pogłębioną znajomością teoretycznych metod obliczeniowych i informatycznych stosowanych do rozwiązywania problemów z chemii.	Student potrafi zastosować pogłębioną wiedzę teoretyczną z zakresu metod obliczeniowych i informatycznych (np. uczenie maszynowe, statystyka, analiza sieci) do analizy dużych zbiorów danych chemicznych. Student potrafi dobrać odpowiednie metody do konkretnego problemu badawczego, przeprowadzić analizę, zinterpretować wyniki oraz ocenić ich wiarygodność.	[SW1] wypowiedź ustna/rozmowa/diskusja [SW5] realizacja zadania problemowego
	[CHEMMU2_U09] Posiada pogłębioną umiejętność przygotowania różnych form prezentacji ustnych w języku polskim i angielskim z chemii.	Student potrafi przygotować i wygłosić różnorodne formy prezentacji ustnych (referat, poster, seminarium, komunikat naukowy) w języku polskim lub angielskim na temat analizy dużych zbiorów danych chemicznych, dostosowując styl i poziom szczegółowości do odbiorców. Student potrafi wykorzystać różnorodne narzędzia i techniki wizualizacji danych, aby przekazać informacje w sposób atrakcyjny i zrozumiały.	[SU1] wypowiedź ustna/rozmowa/diskusja [SU8] obserwacja samodzielnej lub zespołowej pracy studenta

	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[CHEMMU2_U08] Przygotowuje i prezentuje wystąpienia ustne z różnych dziedzin chemii i nauk pokrewnych w języku polskim i angielskim, wykorzystując nabytą wiedzę i umiejętności oraz różnorodne źródła informacji naukowej.	Student potrafi przygotować i wygłosić prezentację ustną w języku polskim lub angielskim na temat analizy dużych zbiorów danych, wykorzystując odpowiednie narzędzia i techniki przetwarzania danych (MapReduce oraz PySpark). Student potrafi krytycznie ocenić wyniki analizy, wyciągnąć wnioski i przedstawić je w sposób zrozumiały dla odbiorców o różnym poziomie wiedzy specjalistycznej.	[SU1] wypowiedź ustna/rozmowa/ dyskusja [SU2] prezentacja/projekt/referat/ raport [SU5] realizacja zadania problemowego
Treści przedmiotu	koncepcja dużych baz danych i BigData, podstawy inżynierii dużych zbiorów danych, infrastruktura sprzętowa BigData (lokalna i chmurowa), algorytm MapReduce wprowadzenie do języka programowania Python, analiza danych w Pythonie, uczenie maszynowe (metody nadzorowane i nienadzorowane), wprowadzenie do Apache Spark i Hadoop konfiguracja środowiska pracy (Python, Spark) i zbiorów danych Big Data, podstawy Spark, odporne rozproszone zbiory danych, przykłady i ćwiczenia RDD, wprowadzenie do SparkSQL, ćwiczenia z komendami SQL, Spark MLlib (regresja liniowa i drzewa decyzyjne z wykorzystaniem Spark ML)		
Wymagania wstępne i dodatkowe			
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	Obserwacja pracy studenta w trakcie zajęć oraz konsultacji	50.0%	30.0%
	Projekt	50.0%	70.0%
Zalecana lista lektur	Podstawowa lista lektur	Apache Spark and PySpark documentation - https://spark.apache.org/docs/ Python documentation - https://docs.python.org/3/ M. Bowles - Machine Learning with Spark and Python®: Essential Techniques for Predictive Analytics	
	Uzupełniająca lista lektur	The Google File System Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung Google	
	Adresy eZasobów	Adresy na platformie eNauczanie:	
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	główne koncepcje Big Data; struktura i właściwości baz danych algorytm MapReduce i jego funkcje mapowania i redukcji; podstawy języka programowania Python (typy, struktury danych, funkcje, biblioteki), podstawowe metody i biblioteki wykorzystywane w analizie danych i uczeniu maszynowym w Pythonie; Apache Spark i Hadoopi jego moduły		
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy		

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.