

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Analiza danych i sieci semantyczne dla Internetu Przedmiotów		Kod 1010512311010510003
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 1
Ścieżka obieralności/specjalność Internet Przedmiotów	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 30 Projekty/seminaria: -		Liczba punktów 5
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) kierunkowy		(ogólnouczelniany, z innego kierunku) z danego kierunku
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne nauki techniczne		Podział ECTS (liczba i %) 5 100% 5 100%
Odpowiedzialny za przedmiot / wykładowca: Odpowiedzialny za przedmiot / wykładowca: dr inż. Tomasz Łukaszewski dr hab. inż. Agnieszka Ławrynowicz email: tomasz.lukaszewski@put.poznan.pl email: agnieszka.lawrynowicz@put.poznan.pl tel. (+48) 61 6652920 tel. (+48) 61 6653026 Wydział Informatyki Wydział Informatyki ul. Piotrowo 3, 60-965 Poznań ul. Piotrowo 3, 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę w zakresie technologii internetowych (w tym języka XML), sztucznej inteligencji i baz danych.
2	Umiejętności:	Student powinien posiadać umiejętność rozwiązywania podstawowych problemów w zakresie projektowania systemów informatycznych i ich realizacji oraz umiejętność pozyskiwania informacji ze wskazanych źródeł. Powinien również rozumieć konieczność poszerzania swoich kompetencji.
3	Kompetencje społeczne	W zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.
Cel przedmiotu:		
1. Przekazanie studentom wiedzy z analizy danych w zakresie statystyki opisowej i uczenia nadzorowanego z przykładów, w tym strumieni danych, wielkich wolumenów danych (ang. BigData). 2. Rozwijanie u studentów umiejętności zastosowania metod uczenia nadzorowanego z wykorzystaniem narzędzia WEKA i bibliotek dla języka Python 3. Przekazanie studentom podstawowej wiedzy w zakresie technologii semantycznych, w tym koncepcji Web 3.0 (semantyczna sieć WWW). 4. Rozwijanie u studentów umiejętności rozwiązywania problemów w zakresie sposobów użytkowania i projektowania systemów wykorzystujących technologie semantyczne.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. ma zaawansowaną i pogłębioną wiedzę w zakresie analizy danych i technologii semantycznych - [K2st_W1] 2. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu sztucznej inteligencji - [K2st_W2] 3. ma zaawansowaną wiedzę szczegółową dotyczącą statystyki opisowej, uczenia nadzorowanego z przykładów, strumieni danych, wielkich wolumenów danych, technologii semantycznych, Web 3.0 - [K2st_W3] 4. ma wiedzę o trendach rozwojowych i nowych osiągnięciach w uczeniu nadzorowanym i technologiach semantycznych - [K2st_W4] 5. zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich z obszaru informatyki dotyczącego uczenia nadzorowanego i technologii semantycznych - [K2st_W6]		
Umiejętności:		

1. potrafi pozyskiwać informacje z literatury, oraz innych źródeł (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K2st_U1]
2. potrafi planować i przeprowadzać eksperymenty w obszarze uczenia nadzorowanego/technologii semantycznych, interpretować uzyskane wyniki i wyciągać wnioski - [K2st_U3]
3. potrafi formułować i testować hipotezy związane z problemami inżynierskimi i prostymi problemami badawczymi w obszarze uczenia nadzorowanego/technologii semantycznych - [K2st_U3]
4. potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody eksperymentalne w obszarze uczenia nadzorowanego/technologii semantycznych - [K2st_U4]
5. potrafi przy rozwiązywaniu zadań integrować wiedzę z różnych obszarów informatyki w szczególności z uczenia nadzorowanego/technologii semantycznych - [K2st_U5]
6. potrafi ocenić przydatność nowych rozwiązań uczenia nadzorowanego/technologii semantycznych w internecie przedmiotów - [K2st_U6]
7. potrafi dokonać krytycznej analizy istniejących rozwiązań technicznych oraz zaproponować ich ulepszenia (usprawnienia) w obszarze uczenia nadzorowanego/technologii semantycznych - [K2st_U8]
8. potrafi ocenić przydatność metod i narzędzi uczenia nadzorowanego/technologii semantycznych w internecie przedmiotów, w szczególności dostrzec ich ograniczenia - [K2st_U9]
9. potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania związane z analizą danych tym zadania nietypowe oraz zadania zawierające komponent badawczy - [K2st_U10]

Kompetencje społeczne:

1. rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st_K1]
2. rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych - [K2st_K2]

Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

a) w zakresie wykładów:

- na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach.

b) w zakresie laboratoriów:

- na podstawie oceny bieżącego postępu realizacji zadań cząstkowych.

Ocena podsumowująca:

a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę wiedzy i umiejętności wykazanych na egzaminie pisemnym o charakterze problemowym. Otrzymanie oceny pozytywnej wymaga uzyskania co najmniej 50% punktów.

b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę prezentowanego przez studenta sprawozdania z realizacji projektu w ramach publicznej prezentacji projektu.

Uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:

- omówienia dodatkowych aspektów zagadnienia,

- efektywność zastosowania zdobytej wiedzy podczas rozwiązywania zadanego problemu,

- uwagi prowadzące do udoskonalenia materiałów dydaktycznych lub procesu dydaktycznego.

Treści programowe

Program wykładu obejmuje następujące zagadnienia: Analiza danych w zakresie uczenia nadzorowanego z przykładów (cel: zrozumienie danych i generalizacji danych na nowe przypadki). Podejścia: drzewa decyzyjne, podejścia regułowe, naiwny klasyfikator Bayesowski. Strumienie danych w Internecie Przedmiotów. Analiza wielkich wolumenów danych (ang. Big Data) z Internetu Przedmiotów. Pojęcie Internetu Semantycznego. Warstwowa architektura języków Internetu Semantycznego. Klasyczne sieci semantyczne i trójkowy model danych. Język opisu zasobów RDF Reprezentacja wiedzy za pomocą ontologii. Język zapytań SPARQL. Pojęcie końcówki SPARQL. Silniki zapytań SPARQL. Modelowanie metadanych. Semantyczne systemy Wiki. Przykłady aktualnych inicjatyw związanych z semantyczną integracją danych i wiedzy (<http://schema.org>, Graf Wiedzy Google), ontologie i schematy metadanych dot. sieci sensorów (W3C SSN) oraz wykorzystanie technologii semantycznych w Internecie Przedmiotów.

Program laboratorium obejmuje następujące zagadnienia: Analiza danych w zakresie statystyki opisowej (miary opisu, rozproszenia) oraz badanie zależności między zmiennymi liczbowymi. Analiza danych (uczenie nadzorowane) z wykorzystaniem narzędzia WEKA i bibliotek dla języka Python (drzewa decyzyjne, systemy regułowe, naiwny klasyfikator Bayesowski, selekcja cech). Reprezentacja danych w modelu RDF. Modelowanie ontologii z wykorzystaniem edytora ontologii (Protégé). Modelowanie metadanych (schema.org, JSON-LD). Przetwarzanie danych semantycznych (repozytoria trójek np. Virtuoso, nierelacyjna baza dokumentów MongoDB). Odpytwanie heterogenicznych źródeł wiedzy za pomocą języka SPARQL. Tworzenie i zarządzanie treścią oraz modelowanie i wyszukiwanie wiedzy w środowisku Semantic Media Wiki.

Literatura podstawowa:		
1. Systemy uczące się. Mirosław Krzyśko, Waldemar Wołyński, Tomasz Górecki, Michał Skorzybut, WNT, 2008.		
2. Eksploracja danych: metody i algorytmy, Tadeusz Morzy, PWN, 2013.		
3. Ontologie w systemach informatycznych, Krzysztof Goczyła, EXIT, 2011.		
Literatura uzupełniająca:		
1. Semantic Web for the Working Ontologist, Dean Allemang and Jim Hendler, Morgan Kaufmann, 2008.		
2. Working with MediaWiki, Yaron Koren, WikiWorks Press, 2012 (aktualizacja 2014).		
3. Semantic data mining. An ontology-based approach. Agnieszka Ławrynowicz, 2017, IOS Pres/AKA Verlag.		
Bilans nakładu pracy przeciętnego studenta		
Czynność		Czas (godz.)
1. udział w wykładach		30
2. przygotowanie do zajęć laboratoryjnych		6
3. udział w zajęciach laboratoryjnych		30
4. dokończenie (w ramach pracy własnej) ćwiczeń laboratoryjnych		10
5. napisanie projektu zaliczeniowego, przygotowanie prezentacji multimedialnej projektu		20
6. udział w konsultacjach związanych z realizacją ćwiczeń laboratoryjnych i projektu		2
7. zapoznanie się ze wskazaną literaturą i materiałami dydaktycznymi		15
8. przygotowanie do egzaminu (10 godz.) i udział w egzaminie (2 godz.)		12
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	125	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	62	2
Zajęcia o charakterze praktycznym	60	2