



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Analiza Danych i Sieci Semantyczne dla Internetu Przedmiotów

Przedmiot

Kierunek studiów

Rok/semestr

Informatyka

1/1

Studia w zakresie (specjalność)

Profil studiów

Internet Przedmiotów

ogólnoakademicki

Poziom studiów

Język oferowanego przedmiotu

drugiego stopnia

polski

Forma studiów

Wymagalność

stacjonarne

obligatoryjny

Liczba godzin

Wykład

Laboratoria

Inne (np. online)

30

30

Ćwiczenia

Projekty/seminaria

Liczba punktów ECTS

5

Wykładowcy

Odpowiedzialny za przedmiot/wykładowca:

Odpowiedzialny za przedmiot/wykładowca:

dr hab. inż. Jerzy Stefanowski, prof. PP

dr hab. inż. Agnieszka Ławrynowicz

dr inż. Tomasz Łukaszewski

Wymagania wstępne

Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę w zakresie technologii internetowych (w tym języka XML), podstaw logiki i baz danych oraz programowania w języku Python. Powinien również posiadać umiejętność pozyskiwania informacji ze wskazanych źródeł oraz mieć gotowość do podjęcia współpracy w ramach zespołu.

Cel przedmiotu

Wprowadzenie do analizy danych w zakresie uczenia maszynowego, w tym klasyfikacji, selekcji cech, strumieni danych i wielkich wolumenów danych (ang. BigData). Przekazanie studentom podstawowej wiedzy w zakresie technologii semantycznych, w tym koncepcji Web 3.0 (semantyczna sieć WWW). Rozwijanie u studentów umiejętności rozwiązywania problemów w zakresie sposobów użytkowania i projektowania systemów wykorzystujących uczenie maszynowe i technologie semantyczne.



Przedmiotowe efekty uczenia się

Wiedza

1. Ma zaawansowaną wiedzę szczegółową dotyczącą klasyfikacji, selekcji cech, strumieni danych, wielkich wolumenów danych, technologii semantycznych, Web 3.0
2. Ma wiedzę o trendach rozwojowych i nowych osiągnięciach w uczeniu maszynowym i technologiach semantycznych
3. Zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich z obszaru informatyki dotyczącego uczenia maszynowego i technologii semantycznych

Umiejętności

1. Potrafi planować i przeprowadzać eksperymenty w obszarze uczenia maszynowego, interpretować uzyskane wyniki i wyciągać wnioski.
2. Potrafi ocenić przydatność metod i narzędzi technologii semantycznych w Internecie Przedmiotów.

Kompetencje społeczne

Rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe.

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Wiedza nabyta w ramach wykładu jest weryfikowana na egzaminie pisemnym. Próg zaliczeniowy: 50% punktów. Zagadnienia zaliczeniowe, na podstawie których opracowywane są pytania zostaną przesłane studentom drogą mailową z wykorzystaniem systemu uczelnianej poczty elektronicznej. Umiejętności nabyte w ramach zajęć laboratoryjnych weryfikowane są podstawie prezentacji będącej wynikiem analizy wskazanego problemu związanego z uczeniem maszynowym i technologiami semantycznymi.

Treści programowe

Program wykładu obejmuje następujące zagadnienia: analiza danych z wykorzystaniem uczenia maszynowego (cel: zrozumienie danych i generalizacji danych na nowe przypadki). Podejścia: drzewa decyzyjne, podejścia regułowe, naiwny klasyfikator Bayesowski, selekcja cech. 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 pogłębienie zagadnień omawianych na wykładach: analiza danych (uczenie maszynowe) z wykorzystaniem modułów dla języka Python (wizualizacja danych, 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). Odpytywanie heterogenicznych źródeł wiedzy za pomocą języka SPARQL. Tworzenie i zarządzanie treścią oraz modelowanie i wyszukiwanie wiedzy w środowisku Semantic Media Wiki.

Metody dydaktyczne

Wykład: prezentacja multimedialna

Ćwiczenia laboratoryjne: ćwiczenia praktyczne, dyskusja, praca w zespole

Literatura

Podstawowa

1. Python: uczenie maszynowe, Sebastian Raschka, Helion 2018
2. Ontologie w systemach informatycznych, Krzysztof Goczyła, EXIT 2011

Uzupełniająca

1. Naczelny Algorytm. Jak jego odkrycie zmieni nasz świat, Pedro Domingos, Helion 2016
2. Semantic Web for the Working Ontologist, Dean Allemang and Jim Hendler, Morgan Kaufmann 2008
3. Working with MediaWiki, Yaron Koren, WikiWorks Press, 2012 (aktualizacja 2014)
4. Semantic data mining. An ontology-based approach. Agnieszka Ławrynowicz. Studies on the Semantic Web, Vol. 29. IOS Pres/AKA Verlag 2017

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	125	5,0
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	65	2,5
Praca własna studenta (studia literaturowe, przygotowanie do egzaminu, dokończenie ćwiczeń laboratoryjnych, wykonanie projektu) ¹	60	2,5

¹ niepotrzebne skreślić lub dopisać inne czynności