



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Analiza Danych i Sieci Semantyczne dla Internetu Przedmiotów

Przedmiot

Kierunek studiów

Informatyka

Studia w zakresie (specjalność)

Internet Przedmiotów

Poziom studiów

drugiego stopnia

Forma studiów

stacjonarne

Rok/semestr

1/1

Profil studiów

ogólnoakademicki

Język oferowanego przedmiotu

polski

Wymagalność

obligatoryjny

Liczba godzin

Wykład

30

Ćwiczenia

Laboratoria

30

Projekty/seminaria

Inne (np. online)

Liczba punktów ECTS

5

Wykładowcy

Odpowiedzialny za przedmiot/wykładowca:

dr hab. inż. Jerzy Stefanowski, prof. PP

dr hab. inż. Agnieszka Ławrynowicz

dr inż. Tomasz Łukaszewski

Odpowiedzialny za przedmiot/wykładowca:

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, analizy skupień, 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, analizy skupień, 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 z zakresu analizy danych obejmuje: klasyfikator najbliższych sąsiadów, drzewa decyzyjne, klasyfikacja Bayesowska, ocena klasyfikatorów, selekcja cech, grupowanie hierarchiczne i k-średnich, zespoły klasyfikatorów, strumienie danych i wielkie wolumeny danych. Z zakresu sieci semantycznych program obejmuje następujące zagadnienia: pojęcie semantycznej sieci WWW i Sieci WWW Rzeczy, warstwowa architektura języków semantycznej sieci WWW, trójkowy model danych, język opisu zasobów RDF, reprezentacja wiedzy za pomocą ontologii, język zapytań SPARQL, dostęp do danych poprzez ontologię, modelowanie metadanych i inżynieria wiedzy, przykłady aktualnych inicjatyw związanych z semantyczną integracją danych i wiedzy (np. <http://schema.org>), 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 z wykorzystaniem modułów dla języka Python: wizualizacja danych, klasyfikator najbliższych sąsiadów, drzewa decyzyjne, naiwny klasyfikator Bayesowski, selekcja cech, grupowanie hierarchiczne i k-średnich. Reprezentacja danych w modelu RDF. Modelowanie ontologii z wykorzystaniem edytora



ontologii (Protégé). Modelowanie metadanych (schema.org, JSON-LD). Przetwarzanie danych semantycznych (repozytoria trójek, MongoDB). Transformacja danych do formatu grafów wiedzy. Odpytywanie heterogenicznych źródeł wiedzy za pomocą języka SPARQL. Dostęp do danych poprzez ontologię (R2RML). Podsumowanie zdobytej wiedzy z technologii semantycznych w ramach mini-projektu.

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
3. Linked Data: Evolving the Web into a Global Data Space (1st edition). Tom Heath and Christian Bizer, Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool, 2011, <http://linkeddatatbook.com/book>

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. Demystifying OWL for the Enterprise, Michael Uschold, Morgan & Claypool Publishers, 2018
4. An Introduction to Ontology Engineering. Keet, C.M. College Publications, volume 20, November 2018
5. Internet rzeczy. Budowa sieci z wykorzystaniem technologii webowych i Raspberry Pi, Dominique Guinard, Vlad Trifa, Helion, 2017
6. 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	2,6
Praca własna studenta (studia literaturowe, przygotowanie do egzaminu, dokończenie ćwiczeń laboratoryjnych, wykonanie projektu) ¹	60	2,4

¹ niepotrzebne skreślić lub dopisać inne czynności