**Nazwa przedmiotu:**

Techniki analizy sieci społecznych

**Koordynator przedmiotu:**

Rajmund Kożuszek

**Status przedmiotu:**

Fakultatywny dowolnego wyboru

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka

**Grupa przedmiotów:**

Przedmioty techniczne - zaawansowane

**Kod przedmiotu:**

TASS

**Semestr nominalny:**

3 / rok ak. 2021/2022

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

1. liczba godzin kontaktowych – 45 godz., w tym
obecność na wykładach 30 godz.,
konsultacje 15 godz.
2. praca własna studenta – 70 godz., w tym
przygotowanie do kolokwium 20 godz.,
realizacja pierwszego projektu 15 godz.,
realizacja drugiego projektu 35 godz.
Łączny nakład pracy studenta wynosi 115 godz., co odpowiada 4 pkt. ECTS.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

1,56 pkt. ECTS, co odpowiada 45 godz. kontaktowym

**Język prowadzenia zajęć:**

polski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

1,74 pkt. ECTS, co odpowiada 50 godz. poświęconym na przygotowanie projektu

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 30h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 0h |
| Projekt:  | 15h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

brak

**Limit liczby studentów:**

30

**Cel przedmiotu:**

Celem przedmiotu jest przedstawienie fenomenu sieci złożonych, rekonstruowanych z śladów cyfrowych, pozostawianych przez ludzi, zwierzęta, maszyny. Kojarzenie takich informacji prowadzi do powstania danych o konkretnej wartości użytkowej. W ramach wykładu przedstawiane są metody opracowane w ramach prowadzonych przez zespół autorski prac badawczych własnych i zespołowych. Wybrane zadania projektowe wymagają przeprowadzenia eksperymentów badawczych dla nowych metod analizy lub nowych zestawów danych.

**Treści kształcenia:**

Wykład:
• Organizacja przedmiotu. Demonstracja przykładowej interaktywnej aplikacji analizującej sieć zrekonstruowaną z danych heterogenicznych, np. będących wynikiem połączenia danych o ruchu drogowym z danymi o gęstości zaludnienia. Podstawowe pojęcia z teorii grafów.
• Pojęcia podstawowe cd. (ścieżka, podgrafy itp.). Miary statystyczne dla grafu. Miary dla wierzchołka (stopień, pośrednictwo, ranking itp.). Algorytmy do wyznaczania w/w miar.
• Specyfika grafów opisujących sieci społeczne – rozkład potęgowy stopni wierzchołków, kręgi i role społeczne w grafie, istotność słabych powiązań, dyfuzja w grafie. Koncepty związane z dynamiką grafu (model Erdősa–Rényiego, powstanie spójnej składowej grafu). Ewolucja sieci: perkolacja i preferencyjne dołączanie, model Barabasi-Albert. Analiza triad, znaczenie słabych więzi.
• Modelowanie rozchodzenia się opinii w sieci, model Isinga, model Friedkina, dyfuzja informacji, masa krytyczna i wypalanie się. Skuteczne i efektywne sterowanie dynamiką stanu węzłów grafu.
• Prezentacja narzędzi do analizy grafów (gotowych aplikacji i modułów dostępnych w różnych językach programowania). Przegląd dostępnych źródeł danych i sposobów tworzenia z nich grafowych reprezentacji powiązań.
• Metody wizualizacji grafów: podstawy teoretyczne i wybrane narzędzia.
• Semantyka powiązań i węzłów – mikroformaty i HTML5. Taksonomie i ontologie.
• Narzędzia i techniki pozyskiwania danych. Przechowywanie pozyskanych danych – bazy nieustrukturyzowane.
• Techniki łączenia danych z różnych źródeł – miary podobieństwa i istotności (Levenshteina, TF-IDF itp.), algorytmy gronowania. Techniki analizy języka naturalnego w zakresie potrzebnym do powiązania dokumentów.
• Język zapytań semantycznych SPARQL i zastosowania praktyczne.
• Grafy dwudzielne i wielodzielne.
• Przykłady praktyczne obejmujące wybór badań wykonanych przez zespół autorski przedmiotu: modelowanie procesu odejść klientów, procesu sprzedaży, rekonstrukcji grafu na podstawie rejestru połączeń telefonicznych.
• Kolokwium.

Projekt:
Studenci realizują dwa zadania projektowe. Celem pierwszego jest zapoznanie się z podstawowymi własnościami sieci złożonych i nabranie wprawy w wykorzystaniu standardowych narzędzi, modułów programistycznych i algorytmów analitycznych. W tym projekcie zarówno dane, jak i narzędzia ich analizy są gotowe, istniejące, wcześniej zbadane. Wyniki projektu są prezentowane w postaci raportu. Drugie zadanie polega na wykonaniu aplikacji łączącej dane z wielu źródeł i dokonującej ich analizy. W tym przypadku zarówno pobranie danych, jak i ich połączenie oraz analiza wymaga albo wytworzenia przez studentów fragmentów specyficznego oprogramowania, albo właściwego skonfigurowania oprogramowania istniejącego. Projekt drugi może mieć formę zadania analitycznego albo aplikacji interaktywnej. Projekt analityczny może mieć charakter badawczy i obejmować weryfikację praktyczną nowych metod analizy albo skonstruowanie własnych metod. Przykładowe źródła danych i aplikacje:
• baza publikacji + baza projektów badawczych: mapa relacji między instytucjami
• profile osób na portalu branżowym + dane statystyczne o migracji: prognoza rozwoju miast
• relacje z wypraw turystycznych w portalu społecznościowym + graf dróg: predykcja tras turystycznych
• dane o wykorzystaniu środków komunikacji miejskiej + mapa ciekawych miejsc: analiza sposobów spędzania wolnego czasu przez mieszkańców
• baza wikipedystów + baza artykułów: mapa zainteresowań ludzkich
• trasy komunikacji miejskiej + informacje o natężeniu ruchu: wyszukiwarka atrakcyjnych miejsc dla inwestycji

**Metody oceny:**

Przewidziano dwa zadania projektowe. Zadanie pierwsze jest realizowane indywidualnie w czasie 3 tygodni, w czasie których studenci wykonują analizę dostarczonych danych przy pomocy istniejących narzędzi. Drugie zadanie wykonywane jest w zespołach dwuosobowych i trwa 5 tygodni. Polega ono na wykonaniu aplikacji łączącej dane z wielu źródeł, przy czym tylko część danych jest dostarczana przez prowadzących, podczas gdy pozostałe uczestnicy projektu muszą wyszukać i pobrać w ogólnie dostępnych serwisach internetowych.

**Egzamin:**

nie

**Literatura:**

J. Wojciechowski, K. Pieńkosz, Grafy i sieci, PWN 2013
A. Fronczak, P. Fronczak Świat sieci złożonych, PWN 2009
M. Russell Mining the Social Web, O’Reilly 2011
M. Tsvetovat, A. Kouznetsov Social Network Analysis for Startups, O’Reilly 2011
M. Kamola, P. Arabas Sieci społeczne i technologiczne, PWN 2018
D. Jurafsky, J. Martin. Speech and language processing. (3rd rev. draft online)
Y. Liu, A. Barabasi. Control Principles of Complex Networks, arXiv.org, 2016
D. Easley, J. Kleinberg. Networks, Crowds, and Markets: Reasoning about a Highly Connected World, OUP, 2010

**Witryna www przedmiotu:**

https://usosweb.usos.pw.edu.pl/kontroler.php?\_action=katalog2/przedmioty/pokazPrzedmiot&prz\_kod=103A-INSID-MSP-TASS

**Uwagi:**

(-)

## Charakterystyki przedmiotowe