

WYDZIAŁ ZARZĄDZANIA**KARTA PRZEDMIOTU****Nazwa przedmiotu w języku polskim:** Pozyskiwanie i analiza danych ze stron www**Nazwa przedmiotu w języku angielskim:** Web scraping and data analysis**Kierunek studiów (jeśli dotyczy):** Inżynieria zarządzania**Specjalność (jeśli dotyczy):** Zarządzanie projektami**Poziom i forma studiów:** II stopień / stacjonarna**Rodzaj przedmiotu:** obowiązkowy**Kod przedmiotu** W08IZZ-SM0011G**Grupa kursów** TAK

	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium
Liczba godzin zajęć zorganizowanych w Uczelni (ZZU)	15		15		
Liczba godzin całkowitego nakładu pracy studenta (CNPS)			75		
Forma zaliczenia			zaliczenie na ocenę		
Dla grupy kursów zaznaczyć kurs końcowy (X)			X		
Liczba punktów ECTS			3		
w tym liczba punktów odpowiadająca zajęciom o charakterze praktycznym (P)			2		
w tym liczba punktów ECTS odpowiadająca zajęciom wymagającym bezpośredniego udziału nauczycieli lub innych osób prowadzących zajęcia (BU)			1,36		

WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I KOMPETENCJI SPOŁECZNYCH

1. Podstawowa znajomość i umiejętność posługiwania się programem R
2. Podstawowa znajomość HTML i CSS

CELE PRZEDMIOTU

C1: Zdobyć wiedzę niezbędną do szybkiego pozyskiwania bardzo dużej ilości danych, automatyzacji tego procesu jak i cyklicznego powtarzania.

C1: Opanowanie umiejętności przetworzenia danych pozyskanych ze stron WWW w użyteczne informacje wspomagające procesy decyzyjne.

C3: Opanowanie umiejętności posługiwania się programem R w całym procesie: od pozyskiwania danych do późniejszej ich analizy.

PRZEDMIOTOWE EFEKTY UCZENIA SIĘ

Z zakresu wiedzy:

PEU_W01: Ma podstawową wiedzę z zakresu pozyskiwania i analizy danych pochodzących ze stron internetowych

Z zakresu umiejętności:

PEU_U01: Potrafi zaprojektować i zaimplementować procedurę pozyskiwania danych ze stron internetowych, a następnie zastosować metody statystyczne do analizy takich danych.

TREŚCI PROGRAMOWE

Forma zajęć - wykład		Liczba godzin
Wy1	Internet jako źródło danych wspomagających procesy decyzyjne.	1
Wy2	Metody przetwarzania danych tekstowych (napisów).	1
Wy3	Wyszukiwanie wzorców, wyrażenia regularne.	2
Wy4	Modele i techniki ekstrakcji danych.	2
Wy5	Parsowanie stron WWW.	3
Wy6	Tworzenie robotów indeksujących. Studium przypadku.	3
Wy7	Ekstrakcja danych poprzez API.	1
Wy8	Kolokwium.	2
	Suma godzin	15

Forma zajęć - laboratorium		Liczba godzin
Lab1	Zasady zaliczenia, regulamin korzystania z laboratorium. R jako środowisko web scraping	1
Lab2	Wybrane operacje na danych, programowanie funkcyjne. Wizualizacja	2
Lab3	Przetwarzanie napisów, wyrażenia regularne	2
Lab4	Omówienie zadania: przetwarzanie napisów na przykładzie wybranej strony WWW	1
Lab4-7	Tworzenie pełzaczy dla wybranego problemu decyzyjnego. Przygotowanie raportu	7
Lab8	Dyskusja i omówienie raportu	2
	Suma godzin	15

STOSOWANE NARZĘDZIA DYDAKTYCZNE

N1. Prezentacja multimedialna

N2. Rozwiązywanie przykładowych problemów, studium przypadku

N3. Pakiet statystyczny R, skrypty R

OCENA OSIĄGNIĘCIA PRZEDMIOTOWYCH EFEKTÓW UCZENIA SIĘ

Oceny (F – formująca (w trakcie semestru), P – podsumowująca (na koniec semestru))	Numer efektu uczenia się	Sposób oceny osiągnięcia efektu uczenia się
F1	PEU_W01	Kolokwium
F2	PEU_U01	Zadanie
F3	PEU_U01	Raport
$P = 0.3 \times F1 + 0.7 \times (0.3 \times F2 + 0.7 \times F3)$		

LITERATURA PODSTAWOWA I UZUPEŁNIAJĄCA**LITERATURA PODSTAWOWA:**

- [1] Kapłon R. *Materiały wykładowe* [dostępne na ePortalu/Teams]
- [2] Mitchell R. *Web Scraping with Python*, 2nd Edition, O'Reilly Media, 2018.
- [3] Wickham H., Çetinkaya-Rundel M., Grolemund G., *R for Data Science*, 2nd Edition, O'Reilly Media, 2023.

LITERATURA UZUPEŁNIAJĄCA:

- [1] Aydin O. *R Web Scraping Quick Start Guide*, Packt Publishing, 2018.
- [2] Fitzgerald M. *Introducing Regular Expressions*, O'Reilly Media, 2012.

OPIEKUN PRZEDMIOTU (IMIE, NAZWISKO, ADRES E-MAIL)

Dr inż. Robert Kapłon; robert.kaplon@pwr.edu.pl