# Politechnika Krakowska im. Tadeusza Kościuszki

# Karta przedmiotu

obowiązuje studentów rozpoczynających studia w roku akademickim 2019/2020

Wydział Informatyki i Telekomunikacji

Kierunek studiów: Informatyka | Profil: Ogólnoakademicki

Forma sudiów: stacjonarne | Kod kierunku: I

Stopień studiów: II

Specjalności: Data science dla inżynierów

## 1 Informacje o przedmiocie

| | |
|---|---|
| Nazwa przedmiotu | Metody pozyskiwania danych i web scarping |
| Nazwa przedmiotu w języku angielskim | Data acquisition and web scraping methods |
| Kod przedmiotu | WIiT I oIIS D10 19/20 |
| Kategoria przedmiotu | Przedmioty specjalnościowe |
| Liczba punktów ECTS | 5.00 |
| Semestry | 1 |

## 2 Rodzaj zajęć, liczba godzin w planie studiów

| Semestr | Wykład | Ćwiczenia | Laboratorium | Laboratorium komputerowe | Seminarium | Projekt |
|---|---|---|---|---|---|---|
| 1 | 30 | 0 | 0 | 0 | 0 | 45 |

## 3 Cele przedmiotu

**Cel 1** Nabycie wiedzy dotyczącej zagadnień związanych z metodami pozyskiwania danych i web scraping.

**Cel 2** Zdobycie praktycznych umiejętności dotyczących zagadnień poruszanych na wykładach, a związanych z metodami pozyskiwania danych i web scraping.

**Cel 3** Wykształcenie praktycznej zdolności do samodzielnego rozwiązywania złożonych problemów poprzez wykorzystanie zdobytej wiedzy.

**Cel 4** Wykształcenie zdolności pracy zespołowej.

## 4 Wymagania wstępne w zakresie wiedzy, umiejętności i innych kompetencji

**1** Zainteresowanie metodami i technikami pozyskiwania danych i web scraping.

**2** Znajomość podstawowych metod tworzenia i eksploatacji systemów informatycznych.

**3** Znajomość systemów operacyjnych i podstaw użytkowania komputerów.

**4** Znajomość języka angielskiego w stopniu umożliwiającym studiowanie literatury fachowej.

## 5 Efekty kształcenia

**EK1 Wiedza** Student ma wiedzę na temat zagadnień związanych z metodami pozyskiwania danych i web scraping.

**EK2 Umiejętności** Student posiada umiejętności dotyczące użycia metod pozyskiwania danych i web scraping.

**EK3 Umiejętności** Student posiada umiejętności samodzielnego rozwiązywania złożonych problemów poprzez wykorzystanie zdobytej wiedzy.

**EK4 Kompetencje społeczne** Student potrafi pracować w zespole.

## 6 Treści programowe

| Wykład | | |
|---|---|---|
| Lp | Tematyka zajęć<br>Opis szczegółowy bloków tematycznych | Liczba godzin |
| **W1** | Znaczenie danych w kontekście uczenia maszynowego. | 2 |
| **W2** | Najpopularniejsze źródła danych wykorzystywane przy nauce data science. | 2 |
| **W3** | Metody przechowywania danych. Relacyjne i nierelacyjne bazy danych. Hurtownie danych. | 4 |
| **W4** | Json, xml, csv, xsl różne formaty zapisu danych. | 2 |
| **W5** | Wyrażenia regularne. | 2 |
| **W6** | Standard XML. XSD. XSL. | 2 |
| **W7** | Wprowadzenie do Xpath. | 2 |
| **W8** | Uzyskiwanie danych z webserwisów typu REST. | 4 |
| **W9** | Uzyskiwanie danych z webserwisów typu SOA. | 4 |
| **W10** | Parsowanie stron internetowych na podstawie biblioteki HtmlAgilityPack. | 2 |
| **W11** | Zaawansowanie parsowanie stron z wykorzystaniem Selenium Web Driver. | 4 |

| Projekt | | |
|---|---|---|
| Lp | Tematyka zajęć<br>Opis szczegółowy bloków tematycznych | Liczba godzin |
| **P1** | Podział na zespoły. Wybór stron internetowych do parsowania. Wybór technologii. Wyznaczenie celu projektu. Przygotowanie harmonogramu prac. | 5 |
| **P2** | Realizacja kolejnych etapów projektu. | 35 |
| **P3** | Końcowe prezentacje projektów. | 5 |

## 7 Narzędzia dydaktyczne

**N1** Wykłady

**N2** Prezentacje multimedialne

**N3** Ćwiczenia projektowe

**N4** Praca w zespole

**N5** Dyskusja

**N6** Konsultacje

## 8 Obciążenie pracą studenta

| Forma aktywności | Średnia liczba godzin na zrealizowanie aktywności |
|---|---|
| **Godziny kontaktowe z nauczycielem akademickim, w tym:** | |
| Godziny wynikające z planu studiów | 75 |
| Konsultacje przedmiotowe | 5 |
| Egzaminy i zaliczenia w sesji | 0 |
| **Godziny bez udziału nauczyciela akademickiego wynikające z nakładu pracy studenta, w tym:** | |
| Przygotowanie się do zajęć, w tym studiowanie zalecanej literatury | 30 |
| Opracowanie wyników | 60 |
| Przygotowanie raportu, projektu, prezentacji, dyskusji | 90 |
| **Sumaryczna liczba godzin dla przedmiotu wynikająca z całego nakładu pracy studenta** | **260** |
| Sumaryczna liczba punktów ECTS dla przedmiotu | 5.00 |

# 9 Sposoby oceny

### Ocena formująca

**F1** Bieżące konsultacje

**F2** Prezentacja wyników projektu

### Ocena podsumowująca

**P1** Projekt zespołowy

### Warunki zaliczenia przedmiotu

**W1** Uzyskanie pozytywnej oceny podsumowującej

## Kryteria oceny

| Efekt kształcenia 1 | |
|---|---|
| Na ocenę 2.0 | Nie ma teoretycznej wiedzy ogólnej w zakresie metod pozyskiwania danych i web scraping. |
| Na ocenę 3.0 | Ma słabą teoretyczną wiedzę ogólną w zakresie metod pozyskiwania danych i web scraping. |
| Na ocenę 3.5 | Ma dostateczną teoretyczną wiedzę ogólną w zakresie metod pozyskiwania danych i web scraping. |
| Na ocenę 4.0 | Ma dobrą teoretyczną wiedzę ogólną w zakresie metod pozyskiwania danych i web scraping. |
| Na ocenę 4.5 | Ma ponad dobrą teoretyczną wiedzę ogólną w zakresie metod pozyskiwania danych i web scraping. |
| Na ocenę 5.0 | Ma bardzo dobrą teoretyczną wiedzę ogólną w zakresie metod pozyskiwania danych i web scraping. |
| Efekt kształcenia 2 | |
| Na ocenę 2.0 | Nie potrafi wykorzystać zdobytej wiedzy celem użycia metod pozyskiwania danych i web scraping. |
| Na ocenę 3.0 | Potrafi słabo wykorzystać zdobytą wiedzę celem użycia metod pozyskiwania danych i web scraping. |
| Na ocenę 3.5 | Potrafi dostatecznie wykorzystać zdobytą wiedzę celem użycia metod pozyskiwania danych i web scraping. |
| Na ocenę 4.0 | Potrafi dobrze wykorzystać zdobytą wiedzę celem użycia metod pozyskiwania danych i web scraping. |
| Na ocenę 4.5 | Potrafi ponad dobrze wykorzystać zdobytą wiedzę celem użycia metod pozyskiwania danych i web scraping. |

| | |
|---|---|
| Na ocenę 5.0 | Potrafi bardzo dobrze wykorzystać zdobytą wiedzę celem użycia metod pozyskiwania danych i web scraping. |
| Efekt kształcenia 3 | |
| Na ocenę 2.0 | Nie potrafi wykorzystać zdobytej wiedzy celem rozwiązania złożonych problemów. |
| Na ocenę 3.0 | Potrafi słabo wykorzystać zdobytą wiedzę celem rozwiązania złożonych problemów. |
| Na ocenę 3.5 | Potrafi dostatecznie wykorzystać zdobytą wiedzę celem rozwiązania złożonych problemów. |
| Na ocenę 4.0 | Potrafi dobrze wykorzystać zdobytą wiedzę celem rozwiązania złożonych problemów. |
| Na ocenę 4.5 | Potrafi ponad dobrze wykorzystać zdobytą wiedzę celem rozwiązania złożonych problemów. |
| Na ocenę 5.0 | Potrafi bardzo dobrze wykorzystać zdobytą wiedzę celem rozwiązania złożonych problemów. |
| Efekt kształcenia 4 | |
| Na ocenę 2.0 | Nie potrafi pracować w zespole. |
| Na ocenę 3.0 | Potrafi słabo pracować w zespole. |
| Na ocenę 3.5 | Potrafi dostatecznie pracować w zespole. |
| Na ocenę 4.0 | Potrafi dobrze pracować w zespole. |
| Na ocenę 4.5 | Potrafi ponad dobrze pracować w zespole. |
| Na ocenę 5.0 | Potrafi bardzo dobrze pracować w zespole. |

## 10 Macierz realizacji przedmiotu

| Efekt kształcenia | Odniesienie danego efektu do szczegółowych efektów zdefiniowanych dla programu | Cele przedmiotu | Treści programowe | Narzędzia dydaktyczne | Sposoby oceny |
|---|---|---|---|---|---|
| EK1 | I2_W02<br>I2_W03<br>I2_W06 | Cel 1 | W1 W2 W3 W4<br>W5 W6 W7 W8<br>W9 W10 W11 | N1 N2 N5 N6 | F1 F2 P1 |

| Efekt kształcenia | Odniesienie danego efektu do szczegółowych efektów zdefiniowanych dla programu | Cele przedmiotu | Treści programowe | Narzędzia dydaktyczne | Sposoby oceny |
|---|---|---|---|---|---|
| EK2 | I2_U03b I2_U07 I2_U08 I2_U12 | Cel 2 | W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 P1 P2 P3 | N1 N2 N3 N4 N5 N6 | F1 F2 P1 |
| EK3 | I2_U03b I2_U07 I2_U08 I2_U12 | Cel 3 | W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 P1 P2 P3 | N1 N2 N3 N4 N5 N6 | F1 F2 P1 |
| EK4 | I2_K02 | Cel 4 | P1 P2 P3 | N3 N4 N5 N6 | F1 F2 P1 |

# 11 Wykaz literatury

# 12 Informacje o nauczycielach akademickich

## Osoba odpowiedzialna za kartę

dr hab. inż. prof.PK. Paweł Pławiak (kontakt: pawel.plawiak@pk.edu.pl)

## Osoby prowadzące przedmiot

**1** dr inż. Paweł Pławiak (kontakt: plawiak@pk.edu.pl)

**2** mgr inż. Wojciech Książek (kontakt: wojtekksiazekk@gmail.com)

**3** mgr inż. Michał Gandor (kontakt: gandor.michal@gmail.com)

**4** mgr inż. Filip Pałka (kontakt: palka.fil@gmail.com)

# 13 Zatwierdzenie karty przedmiotu do realizacji

(miejscowość, data) (odpowiedzialny za przedmiot) (dziekan)

**Przyjmuję do realizacji** (data i podpisy osób prowadzących przedmiot)

. . . . . .

. . . . . .

. . . . . .

. . . . . .