

# POLITECHNIKA KRAKOWSKA IM. TADEUSZA KOŚCIUSZKI

## KARTA PRZEDMIOTU

obowiązuje studentów rozpoczynających studia w roku akademickim 2020/2021

Wydział Informatyki i Telekomunikacji

Kierunek studiów: Informatyka

Profil: Ogólnoakademicki

Forma studiów: stacjonarne

Kod kierunku: I

Stopień studiów: II

Specjalności: Data science

### 1 INFORMACJE O PRZEDMIOCIE

NAZWA PRZEDMIOTU	Przetwarzanie języka naturalnego
NAZWA PRZEDMIOTU W JĘZYKU ANGIELSKIM	Natural Language Processing
KOD PRZEDMIOTU	WiIT I oIIS D9 20/21
KATEGORIA PRZEDMIOTU	Przedmioty specjalnościowe
LICZBA PUNKTÓW ECTS	5.00
SEMESTRY	3

### 2 RODZAJ ZAJĘĆ, LICZBA GODZIN W PLANIE STUDIÓW

SEMESTR	WYKŁAD	ĆWICZENIA	LABORATORIUM	LABORATORIUM KOMPUTERO- WE	SEMINARIUM	PROJEKT
3	30	0	30	0	0	0

### 3 CELE PRZEDMIOTU

**Cel 1** Poznanie zagadnień związanych z przetwarzaniem dokumentów tekstowych.

**Cel 2** Nauka projektowania algorytmów przetwarzania dokumentów tekstowych.

**Cel 3** Praktyka implementacji programów do przetwarzania dokumentów tekstowych.

## 4 WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I INNYCH KOMPETENCJI

- 1 Algorytmy i struktury danych, projektowanie
- 2 Podstawy programowania języków skryptowych (Python)

## 5 EFEKTY KSZTAŁCENIA

**EK1 Kompetencje społeczne** Potrafi pracować indywidualnie i w zespole informatyków, w tym także potrafi zarządzać swoim czasem oraz podejmować zobowiązania i dotrzymywać terminów.

**EK2 Wiedza** Zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu prostych zadań informatycznych z zakresu przetwarzania języka naturalnego.

**EK3 Umiejętności** Umie sformułować specyfikację prostych systemów informatycznych w odniesieniu do sprzętu, oprogramowania systemowego i cech funkcjonalnych aplikacji.

**EK4 Umiejętności** Umie - zgodnie z zadaną specyfikacją - zaprojektować oraz zrealizować prosty system informatyczny, używając właściwych metod, technik i narzędzi.

## 6 TREŚCI PROGRAMOWE

LABORATORIUM		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BŁOKÓW TEMATYCZNYCH	LICZBA GODZIN
L1	Wprowadzenie do języka Python, wyrażenia regularne.	2
L2	Zaawansowane typy danych i ich persystencja.	2
L3	Parsowanie dokumentów tekstowych i analiza leksykalna.	2
L4	Algorytmy pozyskiwania tekstów i budowy korpusu tekstowego.	2
L5	Analiza statystyczna korpusu tekstowego.	2
L6	Indekser dokumentów tekstowych, wyszukiwarka dokumentów.	2
L7	Automaty słownikowe.	2
L8	Własności stop-listy, algorytm budowy.	2
L9	Analiza morfologiczna tekstu, lematyzacja bezkontekstowa.	2
L10	Binaryzacja i wektoryzacja tekstu, wskaźniki dopasowania.	2
L11	Programowanie dynamiczne i wskaźniki podobieństwa.	2
L12	Badanie kontekstu informacji, konstrukcja ngramów.	2
L13	Reprezentacja wiedzy, generowanie tekstu, odpowiedzi.	2
L14	Projekt i implementacja systemu przetwarzania tekstu.	2

LABORATORIUM		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
<b>L15</b>	Testowanie systemu przetwarzania tekstu.	2

WYKŁAD		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
<b>W1</b>	Wprowadzenie do tematyki, podstawowe wiadomości wstępne	2
<b>W2</b>	Narzędzia i języki programowania w przetwarzaniu dokumentów	2
<b>W3</b>	Podstawy gramatyki języka polskiego, formalizacja języka	2
<b>W4</b>	Korpusy tekstowe, Web-crawling, ekstrakcja informacji z tekstu	2
<b>W5</b>	Analiza statystyczna tekstu w języku polskim, słowa stop i kluczowe	2
<b>W6</b>	Indeksowanie dokumentów tekstowych, wyszukiwarki internetowe	2
<b>W7</b>	Analiza leksykalna i algorytmy wyszukiwania wzorca	2
<b>W8</b>	Słowniki, automaty słownikowe, słowniki wielojęzyczne	2
<b>W9</b>	Bezkontekstowa analiza morfologiczna tekstu	2
<b>W10</b>	Kontekst tekstu, technika n-gramów	2
<b>W11</b>	Algorytmy badania podobieństwa dokumentów tekstowych	2
<b>W12</b>	Reprezentacja i przedstawianie wiedzy, generowanie tekstu	2
<b>W13</b>	Podstawy translacji maszynowej tekstu	2
<b>W14</b>	Systemy i aplikacje do przetwarzania języka naturalnego	2
<b>W15</b>	Projektowanie aplikacji do przetwarzania i zarządzania wiedzą	2

## 7 NARZĘDZIA DYDAKTYCZNE

**N1** Wykłady

**N2** Ćwiczenia laboratoryjne

**N3** Praca w grupach

## 8 OBCIĄŻENIE PRACĄ STUDENTA

FORMA AKTYWNOŚCI	ŚREDNIA LICZBA GODZIN NA ZREALIZOWANIE AKTYWNOŚCI
<b>Godziny kontaktowe z nauczycielem akademickim, w tym:</b>	
Godziny wynikające z planu studiów	60
Konsultacje przedmiotowe	16
Egzaminy i zaliczenia w sesji	4
<b>Godziny bez udziału nauczyciela akademickiego wynikające z nakładu pracy studenta, w tym:</b>	
Przygotowanie się do zajęć, w tym studiowanie zalecanej literatury	10
Opracowanie wyników	10
Przygotowanie raportu, projektu, prezentacji, dyskusji	10
Realizacja zadań programistycznych	40
<b>SUMARYCZNA LICZBA GODZIN DLA PRZEDMIOTU WYNIKAJĄCA Z CAŁEGO NAKŁADU PRACY STUDENTA</b>	<b>150</b>
SUMARYCZNA LICZBA PUNKTÓW ECTS DLA PRZEDMIOTU	5.00

## 9 SPOSOBY OCENY

### OCENA FORMUJĄCA

**F1** Ćwiczenie praktyczne

**F2** Sprawozdanie z ćwiczenia laboratoryjnego

**F3** Odpowiedź ustna

### OCENA PODSUMOWUJĄCA

**P1** Średnia ważona ocen formujących

### WARUNKI ZALICZENIA PRZEDMIOTU

**W1** Obecność na zajęciach.

**W2** Aktywność na zajęciach.

### KRYTERIA OCENY

EFEKT KSZTAŁCENIA 1	
NA OCENĘ 2.0	Opanowanie nie więcej niż 50% materiału.

NA OCENĘ 3.0	Opanowanie więcej niż 50% materiału.
NA OCENĘ 3.5	Opanowanie więcej niż 60% materiału.
NA OCENĘ 4.0	Opanowanie więcej niż 70% materiału.
NA OCENĘ 4.5	Opanowanie więcej niż 80% materiału.
NA OCENĘ 5.0	Opanowanie więcej niż 90% materiału.
EFEKT KSZTAŁCENIA 2	
NA OCENĘ 2.0	Opanowanie nie więcej niż 50% materiału.
NA OCENĘ 3.0	Opanowanie więcej niż 50% materiału.
NA OCENĘ 3.5	Opanowanie więcej niż 60% materiału.
NA OCENĘ 4.0	Opanowanie więcej niż 70% materiału.
NA OCENĘ 4.5	Opanowanie więcej niż 80% materiału.
NA OCENĘ 5.0	Opanowanie więcej niż 90% materiału.
EFEKT KSZTAŁCENIA 3	
NA OCENĘ 2.0	Opanowanie nie więcej niż 50% materiału.
NA OCENĘ 3.0	Opanowanie więcej niż 50% materiału.
NA OCENĘ 3.5	Opanowanie więcej niż 60% materiału.
NA OCENĘ 4.0	Opanowanie więcej niż 70% materiału.
NA OCENĘ 4.5	Opanowanie więcej niż 80% materiału.
NA OCENĘ 5.0	Opanowanie więcej niż 90% materiału.
EFEKT KSZTAŁCENIA 4	
NA OCENĘ 2.0	Opanowanie nie więcej niż 50% materiału.
NA OCENĘ 3.0	Opanowanie więcej niż 50% materiału.
NA OCENĘ 3.5	Opanowanie więcej niż 60% materiału.
NA OCENĘ 4.0	Opanowanie więcej niż 70% materiału.
NA OCENĘ 4.5	Opanowanie więcej niż 80% materiału.
NA OCENĘ 5.0	Opanowanie więcej niż 90% materiału.

## 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_U03b	Cel 3	L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 L14 L15	N2 N3	F1 F2 F3
EK2	I2_U07	Cel 1	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15	N1	P1
EK3	I2_U03b	Cel 2	L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 L14 L15	N2 N3	F1 F2 F3
EK4	I2_U12	Cel 3	L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 L14 L15	N2 N3	F1 F2 F3

## 11 WYKAZ LITERATURY

### LITERATURA PODSTAWOWA

- [1] | Lubaszewski W. — *Słowniki komputerowe I automatyczna ekstrakcja informacji z tekstu*, Kraków, 2009, AGH
- [2] | Kłopotek M.A. — *Inteligentne wyszukiwarki internetowe*, Warszawa, 2001, Akademicka Oficyna Wydawnicza Exit
- [3] | Saloni Z., Świdziński M. — *Składnia współczesnego języka polskiego*, Warszawa, 2011, PWN
- [4] | Daciuk J. — *Informacje*: <http://www.eti.pg.gda.pl/jandac/>, Internet, 2011, Politechnika Gdańska
- [5] | Gusfield D. — *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*, Cambridge, 1997, Cambridge University Press
- [6] | Jurafsky D., Martin J.H. — *Speech and language processing, An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, New Jersey, 2000, Prentice-Hall

## 12 INFORMACJE O NAUCZYCIELACH AKADEMICKICH

### OSOBA ODPOWIEDZIALNA ZA KARTĘ

dr inż. Marcin Ciura (kontakt: marcin.ciura@pk.edu.pl)

### OSOBY PROWADZĄCE PRZEDMIOT

1 dr inż. Marcin Ciura (kontakt: marcin.ciura@pk.edu.pl)

## 13 ZATWIERDZENIE KARTY PRZEDMIOTU DO REALIZACJI

---

(miejscowość, data)

(odpowiedzialny za przedmiot)

(dziekan)

PRZYJMUJĘ DO REALIZACJI (data i podpisy osób prowadzących przedmiot)

.....