

POLITECHNIKA KRAKOWSKA IM. TADEUSZA KOŚCIUSZKI

KARTA PRZEDMIOTU

obowiązuje studentów rozpoczynających studia w roku akademickim 2021/2022

Wydział Informatyki i Telekomunikacji

Kierunek studiów: Informatyka

Profil: Ogólnoakademicki

Forma studiów: niestacjonarne

Kod kierunku: I

Stopień studiów: I

Specjalności: Brak specjalności

1 INFORMACJE O PRZEDMIOCIE

NAZWA PRZEDMIOTU	Przetwarzanie języka naturalnego
NAZWA PRZEDMIOTU W JĘZYKU ANGIELSKIM	Natural Language Processing
KOD PRZEDMIOTU	WiIT I oIN D14 21/22
KATEGORIA PRZEDMIOTU	Przedmioty specjalnościowe
LICZBA PUNKTÓW ECTS	5.00
SEMESTRY	8

2 RODZAJ ZAJĘĆ, LICZBA GODZIN W PLANIE STUDIÓW

SEMESTR	WYKŁAD	ĆWICZENIA	LABORATORIUM	LABORATORIUM KOMPUTERO- WE	SEMINARIUM	PROJEKT
8	18	0	18	0	0	0

3 CELE PRZEDMIOTU

Cel 1 Praktyczne poznanie struktury oraz symbolicznych właściwości języka naturalnego. Wykorzystanie w praktyce poznanych metod przetwarzania dla języka naturalnego. Praktyczne wykorzystanie istniejących rozwiązań dla NLP.

Cel 2 Poznanie struktury oraz symbolicznych właściwości języka naturalnego. Algorytmy NLP.

Cel 4 Wykorzystanie w praktyce poznanych metod przetwarzania dla języka naturalnego.

Cel 5 Projektowanie rozwiązań dla zagadnień związanych z NLP.

4 WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I INNYCH KOMPETENCJI

1 Algorytmy i struktury danych, projektowanie.

2 Podstawy programowania języków skryptowych (Python), wyrażenia regularne.

5 EFEKTY KSZTAŁCENIA

EK1 Kompetencje społeczne Student zna i docenia rolę NLP we współczesnym świecie IT

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

EK4 Umiejętności Student umie - zgodnie z zadaną specyfikacją - zaprojektować oraz zrealizować system informatyczny (aplikację) dla zagadnień związanych z NLP. - używając właściwych metod, technik i narzędzi.

EK5 Wiedza Student ma wiedzę na temat metod analizy składniowej, semantycznej oraz wiedzę na temat metod statystycznych dla NLP.

6 TREŚCI PROGRAMOWE

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

WYKŁAD		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
W1	Wprowadzenie do tematyki, podstawowe wiadomości wstępne. Narzędzia i języki programowania w przetwarzaniu dokumentów.	2
W2	Podstawy gramatyki języka polskiego, formalizacja języka	2
W3	Korpusy tekstowe, Web-crawling, ekstrakcja informacji z tekstu	2
W4	Analiza statystyczna tekstu w języku polskim, słowa stop i kluczowe. Indeksowanie dokumentów tekstowych, wyszukiwarki internetowe.	2
W5	Analiza leksykalna i algorytmy wyszukiwania wzorca. Słowniki, automaty słownikowe, słowniki wielojęzyczne.	2
W6	Bezkontekstowa analiza morfologiczna tekstu. Kontekst tekstu, technika n-gramów	2
W7	Algorytmy badania podobieństwa dokumentów tekstowych	2
W8	Reprezentacja i przedstawianie wiedzy, generowanie tekstu. Podstawy translacji maszynowej tekstu	2
W9	Systemy i aplikacje do przetwarzania języka naturalnego. 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
Godziny kontaktowe z nauczycielem akademickim, w tym:	
Godziny wynikające z planu studiów	36
Konsultacje przedmiotowe	16
Egzaminy i zaliczenia w sesji	4
Godziny bez udziału nauczyciela akademickiego wynikające z nakładu pracy studenta, w tym:	
Przygotowanie się do zajęć, w tym studiowanie zalecanej literatury	10
Opracowanie wyników	10
Przygotowanie raportu, projektu, prezentacji, dyskusji	10
Realizacja zadań programistycznych	40
SUMARYCZNA LICZBA GODZIN DLA PRZEDMIOTU WYNIKAJĄCA Z CAŁEGO NAKŁADU PRACY STUDENTA	126
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 obowiązkowych formach zajęć

W2 Aktywność na zajęciach.

KRYTERIA OCENY

EFEKT KSZTAŁCENIA 1	
NA OCENĘ 2.0	Student nie spełnia warunków na ocenę 3.0.

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	Student nie spełnia warunków na ocenę 3.0.
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	Student nie spełnia warunków na ocenę 3.0.
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 5	
NA OCENĘ 2.0	Student nie spełnia warunków na ocenę 3.0.
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	I1_W01 I1_W04 I1_W05 I1_W06 I1_W08 I1_W10 I1_K05 I1_K06	Cel 1	L1 L2 L3 L4 L5 L6 L7 L8 L9	N2 N3	F1 F2 F3
EK2	I1_W08 I1_W09 I1_W11	Cel 1	W1 W2 W3 W4 W5 W6 W7 W8 W9	N1	P1
EK4	I1_W08 I1_W09 I1_W11 I1_W12 I1_U01b I1_U06b I1_U09	Cel 4	L1 L2 L3 L4 L5 L6 L7 L8 L9	N2 N3	F1 F2 F3
EK5	I1_W04 I1_W08 I1_W09 I1_W11 I1_W12	Cel 5	L1	N1	F1

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. Daniel Grzonka (kontakt: daniel.grzonka@pk.edu.pl)

OSOBY PROWADZĄCE PRZEDMIOT

1 mgr inż. Paweł Kisielewicz (kontakt: pkisieles@pk.edu.pl)

13 ZATWIERDZENIE KARTY PRZEDMIOTU DO REALIZACJI

(miejscowość, data)

(odpowiedzialny za przedmiot)

(dziekan)

PRZYJMUJĘ DO REALIZACJI (data i podpisy osób prowadzących przedmiot)

.....