

KARTA KURSU (realizowanego w specjalności)

Data Science

(nazwa specjalności)

Nazwa	Przetwarzanie języka naturalnego
Nazwa w j. ang.	

Koordinator	dr hab. Jozef Kapusta, prof. UKEN	Zespół dydaktyczny
Punktacja ECTS*	st. stacjonarne: 3 st. niestacjonarne: 3	

Opis kursu (cele kształcenia)

Celem kursu jest zapoznanie studentów z podstawowymi metodami przetwarzania języka naturalnego (NLP) oraz ich praktycznym zastosowaniem w analizie tekstu. Studenci poznają klasyczne techniki reprezentacji tekstu, miary podobieństwa dokumentów, metody klasyfikacji i analizy sentymentu, a także sposoby wykorzystania zaawansowanych modeli, takich jak Word2Vec czy GloVe. W ramach kursu omawiane są także elementy morfologii języka, generowania tekstu oraz tłumaczenia maszynowego. Zajęcia mają charakter praktyczny i realizowane są z użyciem języka Python oraz wybranych bibliotek (m.in. NLTK, scikit-learn, Keras). Kurs przygotowuje studentów do rozwiązywania rzeczywistych problemów analizy danych tekstowych z wykorzystaniem metod sztucznej inteligencji.

Warunki wstępne

Wiedza	Student zna podstawy programowania w języku Python oraz podstawowe koncepcje związane z analizą danych i uczeniem maszynowym.
Umiejętności	Student potrafi pracować z bibliotekami Python (np. pandas, numpy, scikit-learn), przetwarzać dane tekstowe oraz tworzyć proste modele klasyfikacyjne.
Kursy	

Efekty uczenia się

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Wiedza	Po zakończeniu kursu student: W01: zna podstawowe techniki przetwarzania języka naturalnego oraz ich zastosowania w analizie tekstu. W02: rozumie sposoby reprezentacji tekstu w przestrzeni wektorowej i metody oceny podobieństwa dokumentów. W03: zna działanie i zastosowanie modeli semantycznych, takich jak Word2Vec czy GloVe. W04: ma wiedzę na temat klasycznych i neuronowych metod przetwarzania tekstu, w tym tłumaczenia maszynowego i generowania języka.	

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Umiejętności	Po zakończeniu kursu student: U01: potrafi przygotować dane tekstowe do analizy i przeprowadzić podstawową eksplorację danych językowych. U02: umie zastosować metody klasyfikacji, analizy sentymentu oraz detekcji anomalii w danych tekstowych. U03: potrafi wykorzystać gotowe narzędzia i biblioteki NLP w języku Python (np. NLTK, Keras). U04: tworzy proste modele analizy i generowania języka z użyciem nowoczesnych metod uczenia maszynowego.	

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Kompetencje społeczne	Po zakończeniu kursu student: K01: potrafi pracować zespołowo nad zadaniami związanymi z przetwarzaniem danych tekstowych i prezentować wyniki analiz. K02: rozumie znaczenie odpowiedzialnego wykorzystywania narzędzi NLP, w tym świadomość możliwych uprzedzeń w danych i modelach.	

Studia stacjonarne

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	10			30			

Studia niestacjonarne

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	5			20			

Opis metod prowadzenia zajęć

Formy sprawdzania efektów uczenia się

	E – learning	Gry dydaktyczne	Ćwiczenia w szkole	Zajęcia terenowe	Praca laboratoryjna	Projekt indywidualny	Projekt grupowy	Udział w dyskusji	Referat	Praca pisemna (esej)	Egzamin ustny	Egzamin pisemny	Zadania problemowe
W01													
W02													
U01													
U02													
K01													

Kryteria oceny	Zaliczenie na ocenę dostateczną otrzymuje student, który potrafi: - Zaliczenie na ocenę dobrą lub bardzo dobrą otrzymuje student, który spełnia warunki oceny dostatecznej, a oprócz tego także:
----------------	---

Uwagi	
-------	--

Treści merytoryczne (wykaz tematów)

1. Wprowadzenie do przetwarzania języka naturalnego (NLP): podstawowe pojęcia i zastosowania; tokenizacja tekstu na poziomie zdań i słów; wyrażenia regularne; wstępna analiza tekstu; stop words; wykorzystanie WordNet. 2. Reprezentacja tekstu w przestrzeni wektorowej: metody reprezentacji słów i dokumentów: Bag-of-Words, TF-IDF. 3. Miary podobieństwa dokumentów: podstawowe metody obliczania podobieństwa między dokumentami – odległość euklidesowa, odległość Jaccarda, podobieństwo cosinusowe. 4. Zaawansowane techniki reprezentacji tekstu: model językowy N-gram, semantyka słów, Word2Vec, GloVe – globalne wektory reprezentacji słów. 5. Zastosowania NLP – analiza sentymentu: wykorzystanie słowników emocji, tworzenie własnych zasobów, budowa klasyfikatora. 6. Zastosowania NLP – klasyfikacja tekstu: przykłady klasyfikacji dokumentów (np. wykrywanie fałszywych wiadomości, analiza wypowiedzi). 7. Analiza składniowa i morfologiczna tekstu: tagowanie części mowy, gramatyka zależności, stemming i lematyzacja, rozpoznawanie nazwanych bytów (NER). 8. Generowanie języka naturalnego: dane sekwencyjne, przetwarzanie wejść w sieciach neuronowych, przykładowe zastosowania (np. generowanie tekstu w stylu autora) z użyciem Pythona i Keras. 9. Tłumaczenie maszynowe: podstawy tłumaczenia statystycznego i neuronowego; miary oceny jakości tłumaczenia: WER, BLEU, METEOR.
--

Wykaz literatury podstawowej

1.

Wykaz literatury uzupełniającej

1.

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) – **studia stacjonarne**

Liczba godzin w kontakcie z prowadzącymi	Wykład	10
	Konwersatorium (ćwiczenia, laboratorium itd.)	30
	Pozostałe godziny kontaktu studenta z prowadzącym	
Liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	
	Realizacja zadań domowych (problemowych) po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca indywidualna lub w grupie)	
	Przygotowanie do egzaminu/zaliczenia	
Ogółem bilans czasu pracy		
Liczba punktów ECTS w zależności od przyjętego przelicznika		3

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) – **studia niestacjonarne**

Liczba godzin w kontakcie z prowadzącymi	Wykład	5
	Konwersatorium (ćwiczenia, laboratorium itd.)	20
	Pozostałe godziny kontaktu studenta z prowadzącym	
Liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	
	Realizacja zadań domowych (problemowych) po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca indywidualna lub w grupie)	
	Przygotowanie do egzaminu/zaliczenia	
Ogółem bilans czasu pracy		
Liczba punktów ECTS w zależności od przyjętego przelicznika		3