

KARTA KURSU

Nazwa	Przetwarzanie języka naturalnego
Nazwa w j. ang.	Natural language processing

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

Opis kursu (cele kształcenia)

This course aims to provide an introduction to the core tasks in natural language processing (NLP). The field of Natural Language Processing (NLP) is primarily concerned with computational models and computer algorithms to process human languages, for example, automatically interpret, generate, and learn natural language. In particular, the topics to be discussed include word embedding, document similarity, syntax, semantics, and their applications in information extraction, machine translation, sentiment analysis, etc.

The course is realized in English.

Warunki wstępne

Wiedza	Znajomość podstaw programowania oraz podstawowych pojęć z zakresu języków formalnych i klasyfikacji wzorców
Umiejętności	Umiejętność programowania proceduralnego i obiektowego. Znajomość jednego z języków: Python, Java, C.
Kursy	

Efekty uczenia się

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Wiedza	After completion of this course, students:	
	W01: understand the significance and importance of text analysis and list some important applications in this area.	K_W01, K_W04, K_W05
	W02: understand several methods used in the field, know their basic characteristics, strengths and weaknesses.	K_W04, K_W05
	W03: understand and distinguish goals in the analysis of word frequency, document similarity and text sentiment.	K_W05
	W04: understand the importance of estimating the quality of machine translation and know the metrics for quality assessment.	K_W05, K_W08

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Umiejętności	After completion of this course, students will be able to:	
	U01: implement basic text mining methods in a selected programming environment.	K_U01, K_U02, K_U03
	U2: implement machine learning algorithms from the natural language processing area for dividing documents into selected categories, based on their text features.	K_U02, K_U03, K_U06, K_U08, K_U09
	U3: design a simple recurrent neural network for a selected natural language generation application.	

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Kompetencje społeczne	After completion of this course, students:	
	K01: works in a team, and takes various roles in it.	K_K01, K_K02
	K02: is aware of the need to share domain knowledge in a way that is understandable to others.	K_K01, K_K05

Studia stacjonarne

Organizacja												
Forma zajęć	Wykład (W)	Ćwiczenia w grupach										
		A		K		L		S		P		E
Liczba godzin	15				30							

Studia niestacjonarne

Organizacja												
Forma zajęć	Wykład (W)	Ćwiczenia w grupach										
		A		K		L		S		P		E
Liczba godzin	10				20							

Opis metod prowadzenia zajęć

W ramach ćwiczeń laboratoryjnych studenci testują i opracowują przykłady programów oraz rozwiązują problemy umożliwiające testowanie poznawanych metod przetwarzania języka naturalnego. Podczas zajęć studenci są zobowiązani osiągnąć wskazane przez prowadzącego rezultaty. Zajęcia, podczas których dochodzi do testowania określonej metody, kończą się weryfikacją poprawności jej implementacji (dokonywaną przez prowadzącego).

W trakcie kursu studenci otrzymują do realizacji poza zajęciami laboratoryjnymi praktyczny projekt grupowy lub indywidualny wymagający kompleksowego podejścia do problematyki implementacji i wykorzystania wybranej metody przetwarzania języka naturalnego.

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	Inne
K_W01					X	X	X	X					X
K_W04					X	X	X	X					X
K_W05					X	X	X						X
K_W08					X	X	X						X
K_U01					X	X	X						X
K_U02					X	X	X						X
K_U03					X	X	X						X
K_U06					X	X	X						X
K_U08					X	X	X						X
K_U09						X	X						
K_K01					X		X	X					
K_K02					X		X						
K_K05						X	X						

Kryteria oceny

Ocena końcowa jest zależna od ocen cząstkowych, systematyczności realizowanych zadań oraz oceny uzyskanej za realizację projektu zespołowego (indywidualnego). W szczególności ocenę dobrą i bardzo dobrą może uzyskać student, który:

- samodzielnie tworzy oprogramowanie wykorzystujące omawiane metody przetwarzania języka naturalnego,
- potrafi zanalizować warunki i obszary stosowalności testowanych algorytmów.

Uwagi

Treści merytoryczne (wykaz tematów)

1. Introduction to Natural Language Processing, Regular expressions, sentence and word tokenization, basic document analyse, stop words, WordNet
2. Word and Document Representations in Vector Space (Bag-of-words, TFI-DF)
3. Document Similarity Measures (Euclidean Distance, Jaccard Distance, Cosine Similarity)
4. Advanced word representations: N-gram language model, language and word meaning, Word2Vec, GloVe - Global Vectors for Word Representation
5. NLP applications: Sentiment analysis - list of positive and negative words, creation of own list, own classifier
6. NLP applications: classification problems (fake news, speech analysis)
7. Morphological parsing: Part-of-Speech tagging, Dependency Grammar, Stemming and lemmatization, Named-entity recognition
8. Python Natural Language Processing (NLP) libraries: NLTK, Stanza, spaCy, Gensim, Polyglot, TextBlob, scikit-learn
9. Natural language generation, sequential data, neural network input mapping, language generation applications, practical examples (name generation, write like Shakespeare, etc.) using Python and Keras
10. Machine Translation: statistical machine translation, neural machine translation, machine translation evaluation, automatic measures (WER, BLEU, Meteor)

Wykaz literatury podstawowej

Wybrane rozdziały:

1. Jurafsky, D., Martin, J.H. (2018). Speech and Language Processing (3rd ed.), Prentice Hall, ISBN-10: 0131873210
2. Clark, A., Fox, Ch., Lappin, S. (2010). The Handbook of Computational Linguistics and Natural Language Processing, Blackwell Publishing Ltd, ISBN:9781405155816

3. Perkins, J. (2014). Python 3 Text Processing with NLTK 3 Cookbook.
4. Aggarwal, Ch. (2018). Machine Learning for Text, Springer. ISBN: 978-3-030-08807-1

Wykaz literatury uzupełniającej

1. Bing, L. (2015). Sentiment Analysis: mining sentiments, opinions, and emotions. Cambridge University Press, 2011
2. Kapusta, J., Drlík, M., Munk, M. (2021). Using of n-grams from morphological tags for fake news classification, In. PeerJ Computer Science 7/2021, s. 1-27, ISSN 2376-5992
1. 3. Kapusta, J., Benko, L., Munková, D., Munk, M. (2021). Analysis of edit operations for post-editing systems, In. International Journal of Computational Intelligence Systems. ISSN 1875-6891

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

liczba godzin w kontakcie z prowadzącymi	Wykład	15
	Konwersatorium (ćwiczenia, laboratorium itd.)	30
	Pozostałe godziny kontaktu studenta z prowadzącym	5
liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	10
	Przygotowanie krótkiej pracy pisemnej lub referatu po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca w grupie)	15
	Przygotowanie do egzaminu/zaliczenia	
Ogółem bilans czasu pracy		75
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	10
	Konwersatorium (ćwiczenia, laboratorium itd.)	20
	Pozostałe godziny kontaktu studenta z prowadzącym	10
liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	10
	Przygotowanie krótkiej pracy pisemnej lub referatu po zapoznaniu się z niezbędną literaturą przedmiotu	
	Przygotowanie projektu lub prezentacji na podany temat (praca w grupie)	15
	Przygotowanie do egzaminu/zaliczenia	5
Ogółem bilans czasu pracy		75
Liczba punktów ECTS w zależności od przyjętego przelicznika		3