**Nazwa przedmiotu:**

Wprowadzenie do przetwarzania języka naturalnego

**Koordynator przedmiotu:**

Rajmund Kożuszek

**Status przedmiotu:**

Fakultatywny dowolnego wyboru

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka

**Grupa przedmiotów:**

Przedmioty techniczne - zaawansowane

**Kod przedmiotu:**

NLP

**Semestr nominalny:**

3 / rok ak. 2021/2022

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

1. liczba godzin kontaktowych – 46 godz.,w tym:
a. obecność na wykładach: 30 godz.,
b. udział w konsultacjach związanych z treścią wykładu: 1 godz.,
c. udział w spotkaniach projektowych: 15 godz.,
2. praca własna studenta – 66 godz., w tym:
a. przygotowanie do wykładów (przejrzenie materiałów z wykładu i dodatkowej literatury, próba rozwiązania ćwiczeń domowych sformułowanych na wykładzie): 6 godz.,
b. realizacja projektu: 10 godz. (zapoznanie się z literaturą i oprogramowaniem) + 30 godz. (wykonanie zadań projektowych) + 10 godz. (sporządzenie dokumentacji) = 50 godz.,
c. przygotowanie do kolokwiów: 10 godz.
Łączny nakład pracy studenta wynosi: 112 godz., co odpowiada 4 pkt. ECTS.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

1,46 pkt. ECTS, co odpowiada 46 godz. kontaktowym

**Język prowadzenia zajęć:**

polski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

1,96 pkt. ECTS, co odpowiada 65 godz. realizacji projektu i spotkań projektowych

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 30h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 0h |
| Projekt:  | 15h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

Biegła umiejętność programowania w przynajmniej jednym z języków: Java, Python, R. Uczestnik posiada podstawową wiedzę na temat uczenia maszynowego ze szczególnym uwzględnieniem sieci neuronowych, zna rachunek prawdopodobieństwa.

**Limit liczby studentów:**

45

**Cel przedmiotu:**

Zapoznanie studentów z metodami automatycznego przetwarzania języka naturalnego.

Treść wykładu obejmuje zagadnienia związane z szeroko rozumianym przetwarzaniem języka naturalnego, ze szczególnym uwzględnieniem przetwarzania języka naturalnego wykorzystywanego w odniesieniu do zasobów sieci internet. W ramach wykładu omówione zostaną podstawowe pojęcia i metody wykorzystywane przy przetwarzaniu dokumentów tekstowych, włącznie z podstawowymi elementami lingwistyki. Omówione zostaną klasyczne metody przetwarzania wykorzystujące reprezentacje dokumentów w postaci przestrzeni wektorowych dokumentów, n-gramów oraz metody wykorzystujące wektory zanurzeń i głębokie sieci neuronowe. Wykorzystanie przedstawionych metod zostanie zaprezentowane w praktycznych zastosowaniach w grupowaniu i klasyfikacji dokumentów, w tłumaczeniu automatycznym, w automatycznej generacji streszczeń i wyszukiwaniu słów kluczowych. Zastosowanie tych metod pokazane zostanie także w odniesieniu do przetwarzania danych pochodzących z sieci internet. Omówione zostaną m.in. algorytmy służące analizie grafu hiperpołączeń tej sieci.

**Treści kształcenia:**

Wykład:

Plan zagadnień poruszanych na wykładach:

1 Wstęp do NLP (3 godz.)
2 Lingwistyczne podstawy, korpusy. Narzędzia lingwistyczne - np. NER, rozbiór zdania, stemmery, lematyzacja. (2 godz.)
3 Reprezentacje dokumentów tekstowych (klasyczne, wykorzystujące wektory zanurzeń - eng. word embeddings) (2 godz.)
4 Glębokie sieci neuronowe w NLP - rekurencyjne sieci neuronowe (Recurrent Neural Networks, RNN, konwolucyjne sieci neuronowe – Convolutional Neural Networks, CNN, model transformera, Recursive Neural Networks, etc.) (3 godz.)
5 Klasyfikacja, grupowanie dokumentów tekstowych (2 godz.)
6 Modele językowe i streszczanie dokumentów (2 godz.)
7 Systemy dialogowe. Moduły: rozumienie języka naturalnego – Natural Language Understanding (NLU), menedżer dialogu – Dialog manager (DM), generowanie języka naturalnego – Natural Language Generation (NLG) (2 godz.)
8 Odpowiadanie na pytania (2 godz.)
9 Ontologie/bazy wiedzy/grafy wiedzy (knowledge graphs) (2 godz.)
10 Tłumaczenie maszynowe (2 godz.)
11 Wyszukiwanie informacji, z uwzględnieniem SEO (2 godz.)
12 Ekstrakcja informacji, z uwzględnieniem ekstrakcji relacji (2 godz.)
13 Aktualności, np. transfer stylu, wykrywanie fałszywych wiadomości, rozpoznawanie emocji (2 godz.)

Projekt:
Celem jest samodzielne zastosowanie metod z obszaru przetwarzania języka naturalnego w praktyce. Projekt realizowany będzie w zespołach 2-osobowych.
Studenci realizować będą zadanie projektowe związane z wymienionymi wyżej zagadnieniami poruszanymi na wykładach.

**Metody oceny:**

Realizacja przedmiotu obejmuje następujące formy zajęć:
● wykład prowadzony w wymiarze 2 godz. tygodniowo,
● projekt realizowany samodzielnie w zespołach,
● konsultacje.
Aktywizacji studentów służą:
● interaktywna formuła wykładu,
● dostarczane po każdym wykładzie ćwiczenia sprawdzające przyswojenie omawianych zagadnień,
● wymóg konsultacji interpretacji tematu i zakresu projektu,
● wymóg przedstawienia do oceny wstępnej dokumentacji projektu,
● wymóg konsultacji zmian interpretacji tematu i zakresu projektu wprowadzanych po ocenie dokumentacji wstępnej.

Sprawdzanie założonych efektów kształcenia realizowane jest przez:
1. ocenę wiedzy i umiejętności wykazanych na sprawdzianie pisemnym,
2. ocenę wiedzy i umiejętności związanych z realizacją zadań projektowych – ocena wykonanych prac implementacyjnych, eksperymentalnych i jakości dokumentacji,
● formatywną ocenę związaną z rozwiązywaniem ćwiczeń domowych formułowanych na wykładzie, udziałem w konsultacjach i interaktywną formą prowadzenia wykładu.

**Egzamin:**

nie

**Literatura:**

Foundations of Statistical Natural Language Processing, Christopher Manning, Hinrich Schütze, 1999
2. Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit, Steven Bird, Ewan Klein, Edward Loper, 2009
3. Handbook of Natural Language Processing and Machine Translation, Olive, Joseph, Christianson, Caitlin, McCary, John (Eds.), 2011
4. Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville, 2016
5. Deep Learning in Natural Language Processing, Li Deng, Yang Liu, 2018
6. NLTK python
7. Spacy library
8. PyTorch
9. Tensorflow

**Witryna www przedmiotu:**

https://usosweb.usos.pw.edu.pl/kontroler.php?\_action=katalog2/przedmioty/pokazPrzedmiot&prz\_kod=103A-INSZI-MSP-NLP

**Uwagi:**

(-)

## Charakterystyki przedmiotowe