



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Przetwarzanie języka naturalnego [N1Inf1>PJN]

Przedmiot

Kierunek studiów
Informatyka

Rok/Semestr
4/7

Studia w zakresie (specjalność)
–

Profil studiów
ogólnoakademicki

Poziom studiów
pierwszego stopnia

Język oferowanego przedmiotu
polski

Forma studiów
niestacjonarne

Wymagalność
obieralny

Liczba godzin

Wykład
12

Laboratorium
12

Inne
0

Ćwiczenia
0

Projekty/seminaria
0

Liczba punktów ECTS

2,00

Koordynatorzy

dr inż. Dawid Wiśniewski
dawid.wisniewski@put.poznan.pl

dr hab. inż. Agnieszka Ławrynowicz prof. PP
agnieszka.lawrynowicz@put.poznan.pl

Wykładowcy

Wymagania wstępne

Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu programowania, podstaw logiki, podstaw sztucznej inteligencji oraz statystyki i analizy danych. Student powinien posiadać umiejętność rozwiązywania podstawowych problemów z zakresu implementacji i oceny kosztu działania prostych algorytmów oraz umiejętność pozyskiwania informacji ze wskazanych źródeł. Student powinien również rozumieć konieczność poszerzania swoich kompetencji i wykazywać gotowość do podjęcia współpracy w ramach zespołu. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.

Cel przedmiotu

1. Przekazanie studentom podstawowej wiedzy na temat przetwarzania języka naturalnego. 2. Rozwijanie u studentów umiejętności rozwiązywania prostych problemów w zakresie sposobów użytkowania i implementacji metod i systemów wykorzystujących przetwarzanie języka naturalnego. 3. Rozwijanie u studentów umiejętności zastosowania metod sztucznej inteligencji i analizy danych do przetwarzania języka naturalnego. 4. Kształtowanie u studentów umiejętności pracy zespołowej.

Przedmiotowe efekty uczenia się

Wiedza:

1. Ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną w zakresie kluczowych zagadnień przetwarzania języka naturalnego, oraz wiedzę szczegółową w zakresie wybranych zagadnień tego obszaru nauki.
2. Ma wiedzę o istotnych kierunkach rozwoju i najważniejszych osiągnięciach w zakresie przetwarzania języka naturalnego oraz innych pokrewnych obszarów jak wykorzystanie metod uczenia maszynowego do celów przetwarzania tekstu i innych związanych z przetwarzaniem języka naturalnego zagadnień sztucznej inteligencji.
3. Zna podstawowe techniki, metody oraz narzędzia wykorzystywane w procesie rozwiązywania zadań związanych z przetwarzaniem języka naturalnego, głównie o charakterze inżynierskim, z zakresu kluczowych zagadnień informatyki.

Umiejętności:

1. Potrafi pozyskiwać informacje z różnych źródeł, w tym z literatury oraz baz danych, zarówno w języku polskim jak i w języku angielskim, właściwie je integrować, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski, oraz wyczerpująco uzasadniać formułowane przez siebie opinie.
2. Potrafi odpowiednio posługiwać się technikami informacyjno-komunikacyjnymi, znajdującymi zastosowanie na różnych etapach realizacji przedsięwzięć informatycznych.
3. Potrafi, formułując i rozwiązując zadania informatyczne, zastosować odpowiednio dobrane metody, w tym metody analityczne, symulacyjne lub eksperymentalne.
4. Potrafi - zgodnie z zadaną specyfikacją - zaprojektować oraz zrealizować urządzenie lub szeroko rozumiany system informatyczny, dobierając język programowania odpowiedni do danego zadania programistycznego oraz używając właściwych metod, technik i narzędzi .
5. Ma umiejętność formułowania algorytmów i ich implementacji z użyciem przynajmniej jednego z popularnych narzędzi.
6. Potrafi planować i realizować proces własnego permanentnego uczenia się oraz zna możliwości dalszego doskonalenia się (studia II i III stopnia, studia podyplomowe, kursy i egzaminy przeprowadzane przez uczelnie, firmy i organizacje zawodowe).

Kompetencje społeczne:

1. Rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe.
2. Ma świadomość znaczenia wiedzy w rozwiązywaniu problemów inżynierskich oraz zna przykłady i rozumie przyczyny wadliwie działających systemów informatycznych, które doprowadziły do poważnej utraty zdrowia, a nawet życia.

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

- a) w zakresie wykładów: na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach (w formie testów) oraz aktywności.
- b) w zakresie laboratoriów: na podstawie oceny bieżącego postępu realizacji zadań cząstkowych.

Ocena podsumowująca:

- a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez: ocenę wiedzy i umiejętności wykazanych w ramach prezentacji będącej wynikiem analizy wskazanego problemu związanego z przetwarzaniem języka naturalnego oraz podsumowanie punktów i omówienie testów z wykładu. Na ostateczną ocenę w zakresie wykładów składają się: punkty z testów dostępnych po wybranych wykładach, punkty z prezentacji oraz punkty za aktywność.
- b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez: ocenę umiejętności związanych z realizacją ćwiczeń laboratoryjnych, ocenę wykonania zadań realizowanych częściowo w trakcie laboratoriów i w części po ich zakończeniu, ocenę projektu jaki studenci będą realizować w celu podsumowania zdobytej wiedzy i umiejętności.

Uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za: omówienia dodatkowych aspektów zagadnienia, wykazanie się ciekawymi umiejętnościami ponadprogramowymi, efektywność zastosowania zdobytej wiedzy podczas rozwiązywania zadanego problemu, uwagi prowadzące do udoskonalenia materiałów dydaktycznych lub procesu dydaktycznego.
Próg zaliczeniowy: 50% punktów.

Treści programowe

W ramach wykładu przedstawiane są następujące zagadnienia:

1. Podstawowa architektura systemów przetwarzania języka naturalnego.
2. Podstawowe techniki przetwarzania tekstu (wyrażenia regularne, filtrowanie wyrazów funkcyjnych, segmentacja, lematyzacja, odległość edycyjna).
3. Klasyfikacja tekstu.
4. Analiza gramatyczna (rozpoznawanie części mowy, tagowanie części mowy, parsery zależnościowe).
5. Statystyczne modelowanie języka, rozproszone reprezentacje i wektory osadzeń: n-gram, word2vec, GloVe, softmax.
6. Reprezentacja wiedzy lingwistycznej w postaci tezaurusów, słowników (WordNet, FrameNet).
7. Zastosowanie deep learning do przetwarzania języka naturalnego (GRU, LSTM, inne modele np. sieci rekursywne).
8. Ekstrakcja informacji i wiedzy z danych tekstowych (wykrywanie encji nazwanych, ekstrakcja relacji).
9. Analiza sentymentu i opinii.
10. Systemy dialogowe, chatboty.

Program laboratorium obejmuje zagadnienia podstawowych metod przetwarzania języka naturalnego z wykorzystaniem popularnych bibliotek i zestawów narzędziowych (np. biblioteka Pythona NLTK, SpaCy, gensim, BeautifulSoup), zagadnienia zastosowania deep learning (głębokich sieci neuronowych) do przetwarzania języka naturalnego (z wykorzystaniem np. TensorFlow/Keras/PyTorch) oraz case studies w zakresie zastosowań poznanych metod (analiza sentymentu i opinii, automatyczne generowanie streszczeń), w szczególności:

1. Szybkie wyszukiwanie/operowanie na tekście - wyrażenia regularne.
2. Klasyfikacja tekstów (tworzenie reprezentacji BagOfWords, Tokenizacja, Lematyzacja, Stemming, normalizacja TF-IDF, SVM, Naiwny Bayes).
3. NGramy (reprezentacja NGram vs BagOfWords, detekcja języka, generowanie tekstu).
4. Embeddingi (wektory osadzone) jako niskowymiarowa alternatywa dla BagOfWords/NGram (podobieństwo w przestrzeni embeddingów, embeddingi do klasyfikacji).
5. Poprawianie literówek z użyciem odległości edycyjnej.
6. Sieci neuronowe (sieci jako sekwencja operacji na macierzach, sieć implementowana bez użycia frameworków).
7. Tworzenie zasobów (Crawling/Scraping danych z sieci, tworzenie zasobów).
8. Sieci rekurencyjne w przetwarzaniu tekstu (RNN od podstaw bez użycia frameworków, idea historii w RNN).
9. Detekcja sentymentu z użyciem zaawansowanych architektur sieci (GRU/LSTM/CNN).
10. Wykrywanie encji nazwanych i fraz rzeczownikowych.
11. Sumaryzacja poprzez wyszukiwanie zdań kluczowych (Key-sentence extraction), modelowanie tematów (topic modelling - LDA).
12. Ekstrakcja informacji (drzewo zależnościowe, rozbiór gramatyczny zdań).

Tematyka zajęć

W ramach wykładu przedstawiane są następujące zagadnienia:

1. Podstawowa architektura systemów przetwarzania języka naturalnego.
2. Podstawowe techniki przetwarzania tekstu (wyrażenia regularne, filtrowanie wyrazów funkcyjnych, segmentacja, lematyzacja, odległość edycyjna).
3. Klasyfikacja tekstu.
4. Analiza gramatyczna (rozpoznawanie części mowy, tagowanie części mowy, parsery zależnościowe).
5. Statystyczne modelowanie języka, rozproszone reprezentacje i wektory osadzeń: n-gram, word2vec, GloVe, softmax.
6. Reprezentacja wiedzy lingwistycznej w postaci tezaurusów, słowników (WordNet, FrameNet).
7. Zastosowanie deep learning do przetwarzania języka naturalnego (GRU, LSTM, inne modele np. sieci rekursywne).
8. Ekstrakcja informacji i wiedzy z danych tekstowych (wykrywanie encji nazwanych, ekstrakcja relacji).

9. Analiza sentymentu i opinii.
10. Systemy dialogowe, chatboty.

Program laboratorium obejmuje zagadnienia podstawowych metod przetwarzania języka naturalnego z wykorzystaniem popularnych bibliotek i zestawów narzędziowych (np. biblioteka Pythona NLTK, SpaCy, gensim, BeautifulSoup), zagadnienia zastosowania deep learning (głębokich sieci neuronowych) do przetwarzania języka naturalnego (z wykorzystaniem np. TensorFlow/Keras/PyTorch) oraz case studies w zakresie zastosowań poznanych metod (analiza sentymentu i opinii, automatyczne generowanie streszczeń), w szczególności:

1. Szybkie wyszukiwanie/operowanie na tekście - wyrażenia regularne.
2. Klasyfikacja tekstów (tworzenie reprezentacji BagOfWords, Tokenizacja, Lematyzacja, Stemming, normalizacja TF-IDF, SVM, Naiwny Bayes).
3. NGramy (reprezentacja NGram vs BagOfWords, detekcja języka, generowanie tekstu).
4. Embeddingi (wektory osadzone) jako niskowymiarowa alternatywa dla BagOfWords/NGram (podobieństwo w przestrzeni embeddingów, embeddingi do klasyfikacji).
5. Poprawianie literówek z użyciem odległości edycyjnej.
6. Sieci neuronowe (sieci jako sekwencja operacji na macierzach, sieć implementowana bez użycia frameworków).
7. Tworzenie zasobów (Crawling/Scraping danych z sieci, tworzenie zasobów).
8. Sieci rekurencyjne w przetwarzaniu tekstu (RNN od podstaw bez użycia frameworków, idea historii w RNN).
9. Detekcja sentymentu z użyciem zaawansowanych architektur sieci (GRU/LSTM/CNN).
10. Wykrywanie encji nazwanych i fraz rzeczownikowych.
11. Sumaryzacja poprzez wyszukiwanie zdań kluczowych (Key-sentence extraction), modelowanie tematów (topic modelling - LDA).
12. Ekstrakcja informacji (drzewo zależnościowe, rozbiór gramatyczny zdań).

Metody dydaktyczne

wykład: prezentacja multimedialna, demonstracja przykładowych rozwiązań, quizy
ćwiczenia laboratoryjne: ćwiczenia praktyczne, dyskusja, praca w zespole, analiza materiałów multimedialnych

Literatura

Podstawowa

1. Natural Language Processing with Python, Steven Bird, Ewan Klein, and Edward Loper, O'Reilly Media, 2009, dostępna online <http://www.nltk.org/book/>

Uzupełniająca

1. Speech and Language Processing (3rd ed. draft), Dan Jurafsky and James H. Martin. Draft chapters in progress, October 16, 2019

4

2. Natural Language Processing in Action. Understanding, analyzing, and generating text with Python, Hobson Lane, Cole Howard, Hannes Hapke, Manning Publications, 2019

3. Foundations of Statistical Natural Language Processing, Chris Manning and Hinrich Schütze, MIT Press. Cambridge, MA: May 1999, <http://nlp.stanford.edu/fsnlp/>

4. The Text Mining Handbook, Ronen Feldman, James Sanger, Cambridge University Press, 2007

5. Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym, Agnieszka Mykowiecka, 2007, Wydawnictwo PJWSTK, Seria: Podręczniki akademickie.

6. Ian Goodfellow, Yoshua Bengio, Aaron Courville "Deep Learning" MIT Press 2016, <http://www.deeplearningbook.org>

7. Zaprojektuj bota. Tworzenie interfejsów konwersacyjnych, Shevat Amir, Helion, 2018

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	50	2,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	24	1,00
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwii/egzaminu, wykonanie projektu)	26	1,00