



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Wybrane tematy w przetwarzaniu języka naturalnego

Przedmiot

Kierunek studiów

Sztuczna Inteligencja

Studia w zakresie (specjalność)

Poziom studiów

Studia drugiego stopnia

Forma studiów

stacjonarna

Rok/semestr

1/2

Profil studiów

ogólnoakademicki

Język oferowanego przedmiotu

angielski

Wymagalność

obligatoryjny

Liczba godzin

Wykład

Laboratoria

Inne (np. online)

Ćwiczenia

Projekty/seminaria

30

Liczba punktów ECTS

2

Wykładowcy

Odpowiedzialny za przedmiot/wykładowca:

dr inż. Mateusz Lango

email: mateusz.lango@cs.put.poznan.pl

Wydział Informatyki i Telekomunikacji

Piotrowo 2, 60-965 Poznań

tel. 61 665 21 24

Odpowiedzialny za przedmiot/wykładowca:

Wymagania wstępne

Student rozpoczynający ten kurs powinien posiadać podstawową wiedzę z zakresu prawdopodobieństwa i statystyki (rozkład normalny, dwumianowy, Dirichleta, Beta i Bernoulliego, estymacja maksymalnej wiarygodności, dywergencja KL, metody Monte Carlo, testowanie hipotez statystycznych), pogłębioną wiedzę z zakresu uczenia maszynowego (zespoły, k-NN, Naive Bayes, SVM,



przekleństwo wymiarowości) oraz uczenia głębokiego (wielowarstwowe sieci neuronowe, rekurencyjne sieci koder-dekoder, sieci konwolucyjne, transformata, backpropagacja, GAN, VAE, strata trójkowa). Dodatkowo zakłada się również średniozaawansowaną wiedzę z zakresu przetwarzania tekstów, równoważną z kursem "Przetwarzanie języka naturalnego" (wyrażenia regularne, stemming, lematyzacja, stopwords, model bag-of-words, miary podobieństwa tekstów, modele BERT/GPT-x/T5, CRF, MEMM, parsowanie zależnościowe, parsowanie konstytutywne, modele tłumaczenia IBM, WordNet itp.)

Student powinien posiadać umiejętność pozyskiwania informacji ze wskazanych źródeł, czytania i rozumienia pracy naukowej, a także przejrzystej prezentacji badań.

W zakresie kompetencji społecznych student musi rozumieć, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe, a także prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.

Cel przedmiotu

Celem kursu jest zapoznanie studentów z metodologią i ostatnimi osiągnięciami badawczymi w zakresie przetwarzania języka naturalnego i technologii języka ludzkiego. Zajęcia koncentrują się na omówieniu najnowszych prac naukowych opublikowanych na czołowych konferencjach takich jak ACL, NAACL, COLING, EMNLP, które dotyczą takich problemów jak tłumaczenie automatyczne, analiza wydzźwięku, systemy dialogowe, generowanie języka naturalnego, odpowiadanie na pytania, analiza składniowa, modelowanie tematów. Dodatkowym celem kursu jest wykształcenie umiejętności krytycznej analizy wyników modeli statystycznych i uczenia maszynowego pod różnymi względami (złożoność obliczeniowa, rodzaj danych treningowych i wymagana wielkość próby, założenia/ograniczenia modelu, metody wnioskowania).

Przedmiotowe efekty uczenia się

Wiedza

1. ma zaawansowaną i pogłębioną wiedzę w zakresie konstrukcji systemów informatycznych przetwarzających język naturalny metodami statystycznymi - [K2st_W3]
2. ma pogłębioną wiedzę o architekturach głębokich sieci neuronowych stosowanych w inżynierii lingwistycznej (w szczególności architektury rekurencyjne i rekursywne) - [K2st_W3]
3. ma zaawansowaną i pogłębioną wiedzę związaną z wybranymi zagadnieniami, takimi jak: modelowanie języka, analiza składniowa, semantyka dystrybucyjna, wykrywanie jednostek nazewniczych, tłumaczenie maszynowe, systemy konwersacyjne - [K2st_W3]
4. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach inżynierii lingwistycznej (w tym nowoczesnych architektur głębokiego uczenia maszynowego) - [K2st_W4]



5. zna zaawansowane metody, techniki i narzędzia stosowane przy budowie systemów dialogowych, translatorów, analizatorów składniowych oraz systemów odpowiadających na pytania - [K2st_W6]

6. zna zaawansowane metody stosowane przy prowadzeniu prac badawczych w zakresie inżynierii lingwistycznej - [K2st_W6]

Umiejętności

1. potrafi pozyskiwać informacje nt. technik inżynierii lingwistycznej z literatury oraz innych źródeł (w języku polskim i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K2st_U1]

2. potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć uczenia maszynowego do rozwiązywania problemów inżynierii lingwistycznej - [K2st_U6]

3. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia - w szczególności w zakresie poznawania nowych technik "state-of-the-art" inżynierii lingwistycznej - [K2st_U16]

Kompetencje społeczne

1. rozumie, że w inżynierii lingwistycznej wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st_K1]

2. rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu inżynierii lingwistycznej i uczenia maszynowego w rozwiązywaniu problemów badawczych i praktycznych - [K2st_K2]

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca w zakresie seminariów:

- na podstawie odpowiedzi na pytania dotyczące prezentowanych artykułów naukowych oraz dyskusji na zajęciach metod/problemów NLP.

Ocena podsumowująca:

a) w zakresie seminariów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę wiedzy i umiejętności wykazanych na teście pisemnym zawierającym pytania otwarte oraz pytania w formie testu wielokrotnego wyboru.

- ocenę przygotowanej przez studenta prezentacji omawiającej wybrane zagadnienia z inżynierii lingwistycznej.

Stosuje się następującą skalę ocen: powyżej 51% punktów - dostateczny, 61% - dostateczny plus, 71% - dobry, 81% - dobry plus, 91% - bardzo dobry

Treści programowe

Kurs wprowadza studenta w prowadzenie badań w obszarze przetwarzania języka naturalnego, wraz z wielością jego problemów i zagadnień takich jak powtarzalność/reprodukowalność wyników badań



naukowych. Podczas seminarium studenci zapoznają się z aktualnymi problemami i trendami badawczymi w NLP, a także z nowoczesnymi zastosowaniami technologii NLP, takimi jak tłumaczenie maszynowe, analiza sentymentu, text-to-speech, rozpoznawanie mowy (ASR), odpowiadanie na pytania, wyszukiwanie informacji, systemy dialogowe i inne. Studenci nauczą się również jak wstępnie ocenić jakość prezentowanych badań naukowych, jak ocenić adekwatność danej pracy naukowej do rozwiązywanego problemu oraz obędą się ze strukturą i językiem używanym w pracach naukowych. Celem seminarium jest również uświadomienie rosnącej roli, jaką technologie NLP odgrywają we współczesnym społeczeństwie oraz związanych z tym kwestii etyki zawodowej.

Metody dydaktyczne

1. Seminarium: prezentacja multimedialna, dyskusja zagadnień i rozwiązań

Literatura

Podstawowa

1. Jurafsky D., Martin J.H.: Speech and Language Processing, III edycja, Pearson/Prentice Hall, 2018 (dostęp online: <https://web.stanford.edu/~jurafsky/slp3/>)
2. Li Deng, Yang Liu: Deep Learning in Natural Language Processing. Springer, 2018 (dostęp poprzez eZasoby biblioteki PP)

Uzupełniająca

1. Goodfellow I., Yoshua B., Courville A.: Deep Learning. Systemy uczące się., PWN, 2018
2. Goldberg, Y., Neural Network Methods in Natural Language Processing, Springer, 2017

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	50	2
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	30	1,5
Praca własna studenta (studia literaturowe, przygotowanie do zajęć – w szczególności prezentacji) ¹	20	0,5

¹ niepotrzebne skreślić lub dopisać inne czynności



POLITECHNIKA POZNAŃSKA

EUROPEJSKI SYSTEM TRANSFERU I AKUMULACJI PUNKTÓW (ECTS)

pl. M. Skłodowskiej-Curie 5, 60-965 Poznań