



UNIWERSYTET
IM. ADAMA MICKIEWICZA
W POZNANIU

Przetwarzanie języka naturalnego Sylabus zajęć

Informacje podstawowe

Kierunek studiów Informatyka Specjalność - Jednostka organizacyjna Wydział Matematyki i Informatyki Poziom studiów studia inżynierskie pierwszego stopnia Forma studiów studia stacjonarne Profil studiów profil ogólnoakademicki		Cykl dydaktyczny 2023/24 Kod zajęć 06INFS.320S.03136.23 Języki wykładowe polski Obligatoryjność Fakultatywny Blok zajęciowy Przedmioty specjalnościowe
Koordynator zajęć	Krzysztof Jassem	
Prowadzący zajęcia	Krzysztof Jassem	
Okres Semestr 6	Forma zajęć / liczba godzin / forma zaliczenia • Wykład: 30, Egzamin • Laboratorium: 30, Zaliczenie z oceną	Liczba punktów ECTS 5

Cele kształcenia dla zajęć

Kod	Cel
C1	Zapoznanie studentów z podstawowymi pojęciami związanymi z przetwarzaniem języka naturalnego
C2	Nauczenie studentów stosowania podstawowych technik przetwarzania języka naturalnego
C3	Zapoznanie studentów z wybranymi metodami uczenia maszynowego stosowanymi w przetwarzaniu języka naturalnego
C4	Nauczenie studentów wybranych metod statystycznych w przetwarzaniu języka naturalnego

Wymagania wstępne

1. Podstawowa umiejętność programowania w języku Python
2. Znajomość systemu operacyjnego Linux

Efekty uczenia się dla zajęć

Kod	Efekty uczenia się dla zajęć w zakresie	Efekty uczenia się dla kierunku	Metody weryfikacji osiągnięcia efektów uczenia się dla zajęć
Wiedzy - Student/ka:			
W1	Zna szeroki wachlarz zastosowań metod przetwarzania języka naturalnego	INF_K3_W08_inz	Egzamin pisemny, Egzamin ustny, Test
W2	Zna matematyczną definicję wyrażeń regularnych; zna ich zastosowania w informatyce	INF_K3_W01	Egzamin pisemny, Egzamin ustny, Test
W3	Zna klasyczne metody analizy leksykalnej i składniowej	INF_K3_W02	Egzamin pisemny, Egzamin ustny, Test
W4	Zna polecenia systemu Linux służące do przetwarzania dokumentów tekstowych	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
W5	Zna metody segmentacji tekstu	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
W6	Zna metody analizy morfologicznej	INF_K3_W02	Egzamin pisemny, Egzamin ustny, Test
W7	Zna popularne zestawy narzędzi przetwarzania języka naturalnego	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
W8	Zna naiwny algorytm klasyfikacji Bayesa i jego zastosowania w przetwarzaniu języka naturalnego	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
W9	Zna pojęcia regresji liniowej i logistycznej oraz ich zastosowania w przetwarzaniu języka naturalnego	INF_K3_W09	Egzamin pisemny, Egzamin ustny, Test
W10	Zna pojęcie sztucznych sieci neuronowych oraz ich zastosowania w przetwarzaniu języka naturalnego	INF_K3_W09	Egzamin pisemny, Egzamin ustny, Test
W11	Zna pojęcie modelu języka	INF_K3_W02	Egzamin pisemny, Egzamin ustny, Test
W12	Zna problemy automatycznej korekty pisowni	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
W13	Zna pojęcie i metody parsowania płytkiego	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
W14	Zna pojęcie i metody parsowania głębokiego	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
W15	Zna podstawowe metody stosowane w tłumaczeniu automatycznym	INF_K3_W03	Egzamin pisemny, Egzamin ustny, Test
Umiejętności - Student/ka:			
U1	Potrafi stosować podstawowe narzędzia do przetwarzania języka naturalnego	INF_K3_U07_inz	Projekt, Zadania programistyczne
U2	Potrafi stosować metody uczenia maszynowego w wybranych zastosowaniach przetwarzania języka naturalnego	INF_K3_U04_inz	Projekt, Zadania programistyczne

Kod	Efekty uczenia się dla zajęć w zakresie	Efekty uczenia się dla kierunku	Metody weryfikacji osiągnięcia efektów uczenia się dla zajęć
U3	Potrafi implementować algorytmy statystyczne stosowane w przetwarzaniu języka naturalnego	INF_K3_U04_inz	Projekt, Zadania programistyczne
U4	Potrafi implementować algorytmy parsowania języka	INF_K3_U05_inz	Projekt, Zadania programistyczne
U5	Potrafi stworzyć prosty system tłumaczenia automatycznego	INF_K3_U05_inz	Projekt, Zadania programistyczne

Treści programowe dla zajęć

Lp.	Treści programowe dla zajęć	Efekty uczenia się dla zajęć	Formy zajęć
1.	Przegląd zastosowań metod przetwarzania języka naturalnego	W1, U1	Wykład, Laboratorium
2.	Wyrażenia regularne i ich zastosowania w przetwarzaniu języka naturalnego	W2, U1	Wykład, Laboratorium
3.	Analiza leksykalna i składniowa z wykorzystaniem pakietu PLY	W3, U1, U4	Wykład, Laboratorium
4.	Przetwarzanie tekstu za pomocą poleceń systemu Linux	W4, U1	Wykład, Laboratorium
5.	Algorytmy segmentacji tekstu	W5, U1	Wykład, Laboratorium
6.	Metody automatycznej analizy leksykalnej i morfologicznej	W6, U1	Wykład, Laboratorium
7.	Zestawy narzędzi przetwarzania języka naturalnego	W7, U1	Wykład, Laboratorium
8.	Naiwny klasyfikator Bayesa i jego zastosowania w przetwarzaniu języka naturalnego	W8, U2	Wykład, Laboratorium
9.	Metody regresji liniowej i logistycznej i ich zastosowania w przetwarzaniu języka naturalnego	W9, U2	Wykład, Laboratorium
10.	Sztuczne sieci neuronowe i ich zastosowania w przetwarzaniu języka naturalnego	W10, U2	Wykład, Laboratorium
11.	Model języka w ujęciu statystycznym	W11, U3	Wykład, Laboratorium
12.	Statystyczne metody automatycznej korekty pisowni	W12, U3	Wykład, Laboratorium
13.	Metody parsowania płytkiego	W13, U4	Wykład, Laboratorium
14.	Metody parsowania głębokiego	W14, U4	Wykład, Laboratorium
15.	Współczesne metody tłumaczenia automatycznego	W15, U5	Wykład, Laboratorium

Informacje dodatkowe

Forma zajęć	Metody i formy prowadzenia zajęć
Wykład	Wykład z prezentacją multimedialną wybranych zagadnień
Laboratorium	Rozwiązywanie zadań (np.: obliczeniowych, artystycznych, praktycznych), Metoda projektu

Forma zajęć	Warunki zaliczenia zajęć
Wykład	Na końcową ocenę składają się wyniki testów przeprowadzanych podczas wykładów (trzy testy rozwiązywane na urządzeniach mobilnych: na początku, w środku i na końcu wykładu) oraz wynik uzyskany na egzaminie pisemnym i ustnym. Skala ocen: 1. bardzo dobry (bdb; 5,0) – od 83% punktów, 2. dobry plus (db plus; 4,5) – od 75% punktów, 3. dobry (db; 4,0) – od 67% punktów, 4. dostateczny plus (dst plus; 3,5) – od 59% punktów, 5. dostateczny (dst; 3,0) – od 50% punktów, 6. niedostateczny (ndst; 2,0) – poniżej 50% punktów.
Laboratorium	Końcowa ocena składa się z sumy punktów uzyskanych za zadania programistyczne i projekty. W pierwszej części kursu studenci rozwiązują zadania programistyczne podane przez prowadzącego i otrzymują punkty za ich poprawne rozwiązywanie. W drugiej części kursu studenci wykonują indywidualne projekty programistyczne związane z tematyką kursu. Projekty punktowane są przez prowadzącego. Skala ocen: 1. bardzo dobry (bdb; 5,0) – od 83% punktów, 2. dobry plus (db plus; 4,5) – od 75% punktów, 3. dobry (db; 4,0) – od 67% punktów, 4. dostateczny plus (dst plus; 3,5) – od 59% punktów, 5. dostateczny (dst; 3,0) – od 50% punktów, 6. niedostateczny (ndst; 2,0) – poniżej 50% punktów.

Literatura

Obowiązkowa

1. Dan Jurafsky and James H. Martin, Speech and Language Processing (3rd ed. draft), dostępne (w dniu 8 marca 2023) na stronie internetowej <https://web.stanford.edu/~jurafsky/slp3/>

Dodatkowa

1. Blog autorstwa Krzysztofa Jassemę omawiający w sposób popularny tematykę wykładu: <https://blog.polengmt.com/blog>

Nakład pracy studenta i punkty ECTS

Rodzaje zajęć studenta	Średnia liczba godzin* przeznaczonych na zrealizowane rodzaje zajęć
Wykład	30
Laboratorium	30
Przygotowanie do zajęć	15
Przygotowanie projektu	40
Czytanie wskazanej literatury	10
Łączny nakład pracy studenta	Liczba godzin 125
Liczba punktów ECTS	ECTS 5

* godzina (lekcyjna) oznacza 45 minut

Efekty uczenia się dla kierunku

Kod	Treść
INF_K3_U04_inz	Absolwent/ka potrafi opracować, przeanalizować, zaprojektować klasyczne algorytmy i systemy informatyczne
INF_K3_U05_inz	Absolwent/ka potrafi pisać, uruchamiać i testować programy w wybranym środowisku programistycznym
INF_K3_U07_inz	Absolwent/ka potrafi zastosować wybrane metody wykorzystywane w wiodących kierunkach badań informatyki
INF_K3_W01	Absolwent/ka zna i rozumie zaawansowane pojęcia z działów matematyki służące do modelowania i rozwiązywania problemów w informatyce
INF_K3_W02	Absolwent/ka zna i rozumie zaawansowane pojęcia i problemy formujące kanon dyscypliny informatyka
INF_K3_W03	Absolwent/ka zna i rozumie narzędzia, technologie i urządzenia informatyczne właściwe dla wybranych obszarów zastosowań oraz podstawy ich działania
INF_K3_W08_inz	Absolwent/ka zna i rozumie zaawansowane pojęcia i problemy związane z zarządzaniem informacją, przetwarzaniem danych i systemami baz danych
INF_K3_W09	Absolwent/ka zna i rozumie zaawansowane zagadnienia związane ze sztuczną inteligencją