

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Reprezentacja wiedzy i przetwarzanie języka naturalnego		Kod 1010514371010510090
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 4 / 7
Ścieżka obieralności/specjalność -	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obieralny
Stopień studiów: I stopień	Forma studiów (stacjonarna/niestacjonarna) niestacjonarna	
Godziny Wykłady: 16 Ćwiczenia: - Laboratoria: 16 Projekty/seminaria: -		Liczba punktów 4
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) kierunkowy z danego kierunku		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne		Podział ECTS (liczba i %) 4 100%
Odpowiedzialny za przedmiot / wykładowca:		
Dr inż. Agnieszka Ławrynowicz email: agnieszka.lawrynowicz@cs.put.poznan.pl tel. 61 6653026 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu programowania (zdobytą na zajęciach z przedmiotów Wprowadzenie do informatyki/Podstawy programowania), podstaw logiki (zdobytą na zajęciach z przedmiotu Logika obliczeniowa), podstaw sztucznej inteligencji (zdobytą na przedmiocie Sztuczna Inteligencja) oraz statystyki i analizy danych (zdobytą na zajęciach z przedmiotu Statystyka i analiza danych).
2	Umiejętności:	Powinien posiadać umiejętność rozwiązywania podstawowych problemów z zakresu implementacji i oceny kosztu działania prostych algorytmów oraz umiejętność pozyskiwania informacji ze wskazanych źródeł.
3	Kompetencje społeczne	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 doboru odpowiednich form reprezentacji wiedzy do rozwiązywanego problemu i ich wykorzystania do modelowania i przetwarzania wiedzy. 4. Rozwijanie u studentów umiejętności zastosowania metod sztucznej inteligencji i analizy danych do przetwarzania języka naturalnego. 5. Kształtowanie u studentów umiejętności pracy zespołowej.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie algorytmów i złożoności, architektury systemów komputerowych, języków i paradygmatów programowania, komunikacji człowiek-komputer, sztucznej inteligencji, baz danych, inżynierii oprogramowania, wspomagania decyzji - [K_W4] 2. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w tematyce przetwarzania języka - [K_W6] 3. zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu prostych zadań informatycznych z zakresu przetwarzania języka naturalnego, komunikacji człowiek-komputer, sztucznej inteligencji, baz danych, inżynierii oprogramowania, wspomagania decyzji - [K_W8]		
Umiejętności:		

1. potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K_U1]
2. potrafi posługiwać się technikami informacyjno-komunikacyjnymi wykorzystywanymi przy realizacji przedsięwzięć informatycznych - [K_U6]
3. ma umiejętność formułowania algorytmów przetwarzania języka naturalnego i ich programowania z użyciem przynajmniej jednego z popularnych narzędzi - [K_U22]

Kompetencje społeczne:

1. rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K_K1]
2. potrafi współdziałać i pracować w grupie, przyjmując w niej różne role - [K_K5]
3. potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania - [K_K6]

Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

- a) w zakresie wykładów:
 - na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach.
- b) w zakresie laboratoriów:
 - na podstawie oceny bieżącego postępu realizacji zadań częściowych.

Ocena podsumowująca:

- a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:
 - ocenę wiedzy i umiejętności wykazanych na zaliczeniu pisemnym o formie testu jednokrotnego wyboru składającego się z ok. 30 pytań, łączna liczba punktów za prawidłowe odpowiedzi : 30, minimalna liczba punktów umożliwiających zaliczenie testu: 16; omówienie wyników testu,
 - podsumowanie punktów i omówienie quizów z wykładu,
 - na ostateczną ocenę w zakresie wykładów składają się: punkty z quizów dostępnych po każdym wykładzie, punkty z testu.
 - 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.
- 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.

Treści programowe

W ramach wykładu przedstawiane są następujące zagadnienia:

- Podstawowa architektura systemów przetwarzania języka naturalnego
- Podstawowe techniki przetwarzania tekstu (wyrażenia regularne, filtrowanie wyrazów funkcyjnych, segmentacja, lematyzacja, odległość edycyjna)
- Klasyfikacja tekstu
- Analiza gramatyczna (rozpoznawanie części mowy, tagowanie POS, parsery zależnościowe)
- Statystyczne modelowanie języka, rozproszone reprezentacje: n-gram, word2vec, GloVe, softmax
- Reprezentacja wiedzy: sieci semantyczne, grafy wiedzy, tezaury, słowniki (WordNet, FrameNet, DBpedia, schema.org, YAGO, inne)
- Ekstrakcja informacji i wiedzy z danych tekstowych (wykrywanie encji nazwanych, ekstrakcja relacji)
- Zastosowanie deep learning do przetwarzania języka naturalnego (GRU, LSTM)
- Analiza sentymentu i opinii
- Automatyczne generowanie streszczeń
- Odpowiadanie na pytania (question answering)
- Systemy dialogowe

Program laboratorium obejmuje zagadnienia podstawowych metod przetwarzania języka naturalnego z wykorzystaniem popularnych bibliotek i zestawów narzędziowych (biblioteka Pythona NLTK), zagadnienia zastosowania deep learning do przetwarzania języka naturalnego (TensorFlow/Theano/Caffe/Keras) oraz case studies w zakresie zastosowań poznanych metod (analiza sentymentu i opinii, automatyczne generowanie streszczeń, odpowiadanie na pytania):

- Modele językowe, tokenizacja, lematyzacja, sentence splitting jako wprowadzenie do przetwarzania języka naturalnego
- Szybkie wyszukiwanie w tekście, wyrażenia regularne
- Wyszukiwanie przybliżone - odległości edycyjne (Levenshteina), uwzględnianie literówek poprzez wyrażenia regularne
- Klasyfikacja tekstu - metoda TF-IDF
- Topic modeling - metoda LDA
- Analiza gramatyczna - parsery zależnościowe i tagowanie POS
- Wykrywanie encji nazwanych, wykrywanie relacji
- Analiza sentymentu i opinii
- Metody generowania streszczeń
- Odpowiadanie na pytania: metody bazujące na wyszukiwaniu informacji (Information Retrieval), metody wykorzystujące słowniki, WordNet oraz metody z wykorzystaniem ekstrakcji relacji, DBpedii, FrameNetu
- Problem rzadkich reprezentacji - word embeddings
- Wprowadzenie do sieci neuronowych (deep learning) w kontekście przetwarzania tekstu (GRU, LSTM)
- Zastosowanie sieci neuronowych (deep learning) do wybranego zadania przetwarzania tekstu

Strona przedmiotu: <https://semantic.cs.put.poznan.pl/nlp>

Metody dydaktyczne:

1. Wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy.
2. Ćwiczenia laboratoryjne: rozwiązywanie zadań, ćwiczenie praktyczne, wykonywanie eksperymentów, dyskusja, praca w zespole, studium przypadków, prezentacja multimedialna, demonstracja.

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/>

Literatura uzupełniająca:

1. Foundations of Statistical Natural Language Processing, Chris Manning and Hinrich Schütze, MIT Press. Cambridge, MA: May 1999, <http://nlp.stanford.edu/fsnlp/>
2. The Text Mining Handbook, Ronen Feldman, James Sanger, Cambridge University Press, 2007
3. Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym, Agnieszka Mykowiecka, 2007, Wydawnictwo PJWSTK, Seria: Podręczniki akademickie.
4. Ian Goodfellow, Yoshua Bengio, Aaron Courville
5. Semantic Web for the Working Ontologist, Dean Allemang and Jim Hendler, Morgan Kaufmann, 2008
6. Linked Data: Evolving the Web into a Global Data Space (1st edition), Tom Heath and Christian Bizer, Synthesis Lectures on the Semantic Web: Theory and Technology, 2011, 1:1, 1-136. Morgan & Claypool, dostępna online <http://linkeddatatool.com>
7. Voice Application Development for Android, Michael F. McTear, Zoraida Callejas, PACKT Publishing, 2013

Bilans nakładu pracy przeciętnego studenta		
Czynność	Czas (godz.)	
1. udział w zajęciach laboratoryjnych :	16	
2. przygotowanie do ćwiczeń laboratoryjnych:	14	
3. dokończenie (w ramach pracy własnej) zadań z ćwiczeń laboratoryjnych:	22	
4. udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności ćwiczeń laboratoryjnych / projektu	2 5	
5. rozwiązanie quizu z wykładu	14	
6. przygotowanie do testu	16	
7. udział w wykładach	10	
8. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.), 100 stron		
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	99	4
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	34	1
Zajęcia o charakterze praktycznym	52	2