

<b>KARTA OPISU MODUŁU KSZTAŁCENIA</b>		
Nazwa modułu/przedmiotu <b>Przetwarzanie języka naturalnego</b>		Kod <b>1010511361010510144</b>
Kierunek studiów <b>Informatyka</b>	Profil kształcenia (ogólnoakademicki, praktyczny) <b>ogólnoakademicki</b>	Rok / Semestr <b>3 / 6</b>
Ścieżka obieralności/specjalność <b>-</b>	Przedmiot oferowany w języku: <b>polski</b>	Kurs (obligatoryjny/obieralny) <b>obieralny</b>
Stopień studiów: <b>I stopień</b>	Forma studiów (stacjonarna/niestacjonarna) <b>stacjonarna</b>	
Godziny Wykłady: <b>30</b> Ćwiczenia: <b>-</b> Laboratoria: <b>30</b> Projekty/seminaria: <b>-</b>		Liczba punktów <b>4</b>
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) <b>kierunkowy</b>		(ogólnouczelniany, z innego kierunku) <b>z danego kierunku</b>
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki <b>nauki techniczne</b>  <b>nauki techniczne</b>		Podział ECTS (liczba i %) <b>4 100%</b>  <b>4 100%</b>
Odpowiedzialny za przedmiot / wykładowca: <b>Odpowiedzialny za przedmiot / wykładowca:</b>		
dr hab. inż. Agnieszka Ławrynowicz email: alawrynowicz@cs.put.poznan.pl tel. (+48) 61 6653026 Wydział Informatyki ul. Piotrowo 3, 60-965 Poznań		mgr inż. Dawid Wiśniewski email: dwisniewski@cs.put.poznan.pl tel. (+48) 61 6653053 Wydział Informatyki ul. Piotrowo 3, 60-965 Poznań
<b>Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:</b>		
1	<b>Wiedza:</b>	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	<b>Umiejętności:</b>	Student powinien posiadać umiejętność rozwiązywania podstawowych problemów z zakresu implementacji i oceny kosztu działania prostych algorytmów oraz umiejętność pozyskiwania informacji ze wskazanych źródeł.
3	<b>Kompetencje społeczne</b>	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.
<b>Cel przedmiotu:</b>		
<ol style="list-style-type: none"> <li>Przekazanie studentom podstawowej wiedzy na temat przetwarzania języka naturalnego.</li> <li>Rozwijanie u studentów umiejętności rozwiązywania prostych problemów w zakresie sposobów użytkowania i implementacji metod i systemów wykorzystujących przetwarzanie języka naturalnego.</li> <li>Rozwijanie u studentów umiejętności doboru odpowiednich form reprezentacji wiedzy do rozwiązywanego problemu i ich wykorzystania do modelowania i przetwarzania wiedzy.</li> <li>Rozwijanie u studentów umiejętności zastosowania metod sztucznej inteligencji i analizy danych do przetwarzania języka naturalnego.</li> <li>Kształtowanie u studentów umiejętności pracy zespołowej.</li> </ol>		
<b>Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia</b>		
<b>Wiedza:</b>		

<p>1. ma rozszerzoną i pogłębioną wiedzę z matematyki przydatną do formułowania i rozwiązywania złożonych zadań informatycznych dotyczących m.in. programowania w logice, rachunku prawdopodobieństwa, formalnej specyfikacji i weryfikacji oprogramowania - [K1st_W1]</p> <p>2. ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną w zakresie kluczowych zagadnień z zakresu przetwarzania języka naturalnego, oraz wiedzę szczegółową z tego zakresu - [K1st_W4]</p> <p>3. ma wiedzę o istotnych kierunkach rozwoju i najważniejszych osiągnięciach informatyki z zakresu przetwarzania języka naturalnego - [K1st_W5]</p> <p>4. ma podstawową wiedzę o cyklu życia systemów informatycznych przetwarzania języka naturalnego, zarówno sprzętowych jak i programowych, a w szczególności o zachodzących w nich kluczowych procesach - [K1st_W6]</p> <p>5. zna podstawowe techniki, metody oraz narzędzia wykorzystywane w procesie rozwiązywania zadań informatycznych z zakresu przetwarzania języka naturalnego - [K1st_W7]</p>
<p><b>Umiejętności:</b></p> <p>1. potrafi pozyskiwać informacje z zakresu przetwarzania języka naturalnego 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 - [K1st_U1]</p> <p>2. potrafi odpowiednio posługiwać się technikami informacyjno-komunikacyjnymi, znajdującymi zastosowanie na różnych etapach realizacji przedsięwzięć informatycznych z zakresu przetwarzania języka naturalnego - [K1st_U2]</p> <p>3. potrafi, formułując i rozwiązując zadania informatyczne z zakresu przetwarzania języka naturalnego, zastosować odpowiednio dobrane metody, w tym metody analityczne lub eksperymentalne - [K1st_U4]</p> <p>4. potrafi zaprojektować oraz zrealizować system informatyczny z zakresu przetwarzania języka naturalnego, dobierając odpowiedni język programowania oraz używając właściwych metod, technik i narzędzi - [K1st_U10]</p> <p>5. ma umiejętność formułowania algorytmów i ich implementacji z użyciem przynajmniej jednego z popularnych narzędzi - [K1st_U11]</p> <p>6. potrafi planować i realizować proces własnego permanentnego uczenia się z przetwarzania języka naturalnego - [K1st_U19]</p>
<p><b>Kompetencje społeczne:</b></p> <p>1. rozumie, że w informatyce wiedza i umiejętności z zakresu przetwarzania języka naturalnego bardzo szybko stają się przestarzałe - [K1st_K1]</p> <p>2. ma świadomość znaczenia wiedzy w rozwiązywaniu problemów inżynierskich z zakresu przetwarzania języka naturalnego - [K1st_K2]</p> <p>3. potrafi myśleć i działać w sposób przedsiębiorczy, m.in. znajdując komercyjne zastosowania dla tworzonego oprogramowania, mając na uwadze nie tylko korzyści biznesowe, ale również społeczne prowadzonej działalności - [K1st_K3]</p>

### Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

- a) w zakresie wykładów:
- ? na podstawie aktywnego uczestnictwa w wykładach.
  - ? na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach (quizz)
- b) w zakresie laboratoriów:
- ? na podstawie oceny bieżącego postępu realizacji zadań cząstkowych,
  - ? na podstawie postępów projektu.

Ocena podsumowująca:

- a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:
- ? podsumowanie punktów i omówienie ?quizzów? z wykładu,
  - ? prezentacje studentek i studentów, na wybrane tematy związane z przetwarzaniem języka naturalnego,
  - ? na ostateczną ocenę w zakresie wykładów składają się: punkty z ?quizzów? dostępnych po każdym wykładzie, punkty za aktywność, punkty za prezentację.
- 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.

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

#### 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/>
2. Speech and Language Processing (3rd ed. draft), Dan Jurafsky and James H. Martin. Draft chapters in progress, August 28, 2017 (dostępna online)

#### 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. Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press 2016, <http://www.deeplearningbook.org>
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://linkeddatadatabook.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	30
2. przygotowanie do ćwiczeń laboratoryjnych	3
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	4
6. przygotowanie do testu	30
7. udział w wykładach	4
8. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.), 40 stron	
<b>Obciążenie pracą studenta</b>	
<b>forma aktywności</b>	<b>godzin</b>
<b>ECTS</b>	
Łączny nakład pracy	100
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	62
Zajęcia o charakterze praktycznym	55