

OPISU MODUŁU KSZTAŁCENIA (SYLABUS)

I Informacje ogólne

1. Laboratorium: Automatyczne przetwarzanie tekstu
2. Kod modułu kształcenia: 08-KODM-LABPT4
3. Rodzaj modułu kształcenia: fakultatywny
4. Kierunek studiów: kognitywistyka
5. Poziom studiów: jednolite studia magisterskie
6. Rok studiów (jeśli obowiązuje): V
7. Semestr: letni
8. Rodzaje zajęć i liczba godzin: 45h ćw
9. Liczba punktów ECTS: 8
10. Imię, nazwisko, tytuł/stopień naukowy, adres e-mail wykładowcy (wykładowców) / prowadzących zajęcia: dr Barbara Konat (bkonat@amu.edu.pl)
11. Język wykładowy: polski

II Informacje szczegółowe

1. Cel (cele) modułu kształcenia

Laboratorium przygotowuje studentów do podejmowania roli analityka danych językowych (ang. data analyst) w zespołach badawczych oraz w projektach komercyjnych. Obejmuje dziedzinę informatyki, jaką jest inżynieria lingwistyczna (language engineering) oraz przetwarzanie języka naturalnego (natural language processing, NLP), szczególnie zaś automatyczną analizę danych językowych oraz komunikację człowiek-komputer.

C1	Przekazanie wiedzy w zakresie metod i technik przetwarzania języka naturalnego.
C2	Rozwinięcie umiejętności posługiwania się językami programowania.

2. Wymagania wstępne w zakresie wiedzy, umiejętności oraz kompetencji społecznych (jeśli obowiązują)

Podstawowa wiedza z zakresu językoznawstwa ogólnego, znajomość języka angielskiego na poziomie umożliwiającym korzystanie z anglojęzycznych źródeł literaturowych.

3. Efekty kształcenia w zakresie wiedzy, umiejętności oraz kompetencji społecznych dla modułu kształcenia i odniesienie do efektów kształcenia dla kierunku studiów

Symbol efektów kształcenia	Po zakończeniu modułu (przedmiotu) i potwierdzeniu osiągnięcia efektów kształcenia student potrafi:	Odniesienie do efektów kształcenia dla kierunku studiów
LABPT4_01	zna i rozumie etapy procesu badawczego w przetwarzaniu języka naturalnego	K_W08
LABPT4_02	stworzyć celowy korpus tekstów (poprzez właściwy dobór materiału oraz metodyki anotacji)	K_W10
LABPT4_03	przeprowadzić anotację korpusu tekstów z użyciem dostępnych narzędzi oraz obliczyć zgodność między anotacjami (IAA – inter annotator agreement)	K_W08, K_W10
LABPT4_04	napisać prosty program komputerowy pozwalający na analizę zaanotowanego korpusu	K_W13, K_U01, K_U10
LABPT4_05	korzysta ze źródeł literaturowych, głównie w języku angielskim	K_U01
LABPT4_06	zna i prawidłowo posługuje się podstawowymi pojęciami z zakresu językoznawstwa obliczeniowego	K_W08
LABPT4_07	wyciąga wnioski i porządkuje efekty dyskusji	K_K02

4. Treści kształcenia

Nazwa modułu kształcenia: Automatyczne przetwarzanie tekstu		
Symbol treści kształcenia	Opis treści kształcenia	Odniesienie do efektów kształcenia modułu
TK_01	Elementy procesu przetwarzania języka naturalnego	LABPT4_01, LABPT4_05, LABPT4_06, LABPT4_07
TK_02	Podstawowe narzędzia przetwarzania języka naturalnego (NLTK Python, ONLP Java lub inne narzędzia/pakiety)	LABPT4_01, LABPT4_05, LABPT4_06, LABPT4_07
TK_03	Przygotowanie prostego korpusu z dwuwartościową anotacją (np. korpus tweetów zaanotowanych wydźwiękiem pozytywnym i negatywnym)	LABPT4_02
TK_04	Trening prostego (dzwuwartosciowego) modelu automatycznego klasyfikatora tekstu z uczeniem maszynowym z gotowym zestawem cech (np. bag of words lub n-gram)	LABPT4_04
TK_05	Ewaluacja prostego modelu automatycznego klasyfikatora tekstu (metoda F1-score)	LABPT4_03
TK_06	Etapy tworzenia złożonych algorytmów analizy dyskursu (np. Rhetorical Structure Theory, teoria argumentacji)	LABPT4_01, LABPT4_05, LABPT4_06, LABPT4_07
TK_07	Narzędzia anotacji dla analizy dyskursu	LABPT4_01, LABPT4_05, LABPT4_06, LABPT4_07
TK_08	Tworzenie korpusu dla analizy dyskursu	LABPT4_03
TK_09	Trening złożonego modelu automatycznej analizy dyskursu	LABPT4_04
TK_10	Ewaluacja złożonego modelu automatycznej analizy dyskursu (metoda F1-score lub inna metoda)	LABPT4_04

5. Zalecana literatura

Ingersoll, Grant S., Thomas S. Morton, and Andrew L. Farris. *Taming text: how to find, organize, and manipulate it*. Manning Publications Co., 2013.

Apache OpenNLP Developer Documentation

<https://opennlp.apache.org/documentation/1.7.0/manual/opennlp.html> (Dostęp 16.01.2017)

Natural Language Processing with Python – NLTK <http://www.nltk.org/book/> (Dostęp 16.01.2017)

Stede, Manfred. "Discourse processing." *Synthesis Lectures on Human Language Technologies* 4.3 (2011).

6. Informacja o przewidywanej możliwości wykorzystania e-learningu: brak możliwości

7. Informacja o tym, gdzie można zapoznać się z materiałami do zajęć, instrukcjami do laboratorium, itp.

<https://pan-pl.academia.edu/BarbaraKonat>

III Informacje dodatkowe

1. Odniesienie efektów kształcenia i treści kształcenia do sposobów prowadzenia zajęć i metod oceniania

Nazwa modułu (przedmiotu): Automatyczne przetwarzanie tekstu			
Symbol efektu kształcenia dla modułu	Symbol treści kształcenia realizowanych w trakcie zajęć	Sposoby prowadzenia zajęć umożliwiające osiągnięcie założonych efektów kształcenia	Metody oceniania stopnia osiągnięcia założonego efektu kształcenia
LABPT4_01, LABPT4_02, LABPT4_03	TK_01-TK_11	Laboratorium komputerowe, podczas którego studenci samodzielnie tworzą dane do obróbki statystycznej i programy komputerowe.	Bieżąca ocena indywidualnej pracy studentów podczas zajęć.
LABPT4_01, LABPT4_02, LABPT4_03	TK_01-TK_11	Dyskusja podczas omawiania zagadnień teoretycznych i metodologicznych.	Bieżąca ocena przygotowania studentów i ich udziału w dyskusji podczas zajęć.
LABPT4_01, LABPT4_02, LABPT4_03, LABPT4_04	TK_01-TK_11	Praca samodzielna – przygotowanie raportu z wynikami badań w strukturze odpowiedniej dla artykułu naukowego (metody, materiał, wyniki).	Ocena raportu pod kątem zgodności z wymaganiami.
LABPT4_05	TK_01-TK_11	Praca samodzielna – przygotowanie plików dołączanych do raportów (pliki zbiorów danych, skrypty).	Ocena poprawności przygotowanych plików.

2. Obciążenie pracą studenta (punkty ECTS)

Nazwa modułu (przedmiotu):	
Forma aktywności	Średnia liczba godzin na zrealizowanie aktywności
Godziny zajęć (wg planu studiów) z nauczycielem	45
Praca własna studenta: przygotowanie raportów	70
Praca własna studenta: analiza literatury	25
Praca własna studenta: przygotowanie do zajęć	60
SUMA GODZIN	200
SUMARYCZNA LICZBA PUNKTÓW ECTS DLA MODUŁU (PRZEDMIOTU)	8

3. Sumaryczne wskaźniki ilościowe

- Liczba punktów ECTS, którą student uzyskuje na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich: 2
- Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym, takich jak zajęcia laboratoryjne i projektowe: 8

4. Kryteria oceniania

Obowiązek 80% obecności.

Bieżąca aktywność na zajęciach, dyskusja oraz przygotowywanie efektów bieżących zajęć (bazy danych, skrypty) – 20 punktów.

Dwa raporty z wykonanych zadań, z dołączonymi danymi i skryptami – 20 punktów każdy.

Niezapowiedziana kartkówka – 10 punktów.

Ocena końcowa wyznaczona jest następującą skalą:

powyżej 63 do 70 pkt: bdb

powyżej 56 do 63 pkt: db+

powyżej 49 do 56 pkt: db

powyżej 42 do 49 pkt: dst+

powyżej 35 do 42 pkt: dst

do 35 pkt: ndst