

Eksploracja zasobów internetowych - opis przedmiotu

Informacje ogólne	
Nazwa przedmiotu	Eksploracja zasobów internetowych
Kod przedmiotu	11.3-WE-BEP-EZI
Wydział	Wydział Informatyki, Elektrotechniki i Automatyki
Kierunek	Biznes elektroniczny
Profil	praktyczny
Rodzaj studiów	pierwszego stopnia z tyt. inżyniera
Semestr rozpoczęcia	semestr zimowy 2022/2023

Informacje o przedmiocie	
Semestr	5
Liczba punktów ECTS do zdobycia	4
Typ przedmiotu	obowiązkowy
Język nauczania	polski
Sylabus opracował	• dr hab. inż. Artur Gramacki, prof. UZ

Formy zajęć					
Forma zajęć	Liczba godzin w semestrze (stacjonarne)	Liczba godzin w tygodniu (stacjonarne)	Liczba godzin w semestrze (niestacjonarne)	Liczba godzin w tygodniu (niestacjonarne)	Forma zaliczenia
Wykład	15	1	9	0,6	Zaliczenie na ocenę
Projekt	30	2	18	1,2	Zaliczenie na ocenę

Cel przedmiotu

1. Zapoznanie studentów z podstawowymi modelami i technikami odkrywania informacji znajdujących się w sieci Internet.
2. Zapoznanie ze sposobami działania algorytmów text mining.
3. Ukształtowanie umiejętności eksploracji zasobów internetowych w oparciu o oprogramowanie statystyczne.

Wymagania wstępne

podstawy analityki biznesowej

Zakres tematyczny

Rodzaje informacji w internecie. Wprowadzenie do tematyki Text Mining. Przeszukiwanie informacji tekstowych. Wstępne przetwarzanie dokumentów tekstowych: usuwanie zbędnych elementów z dokumentów tekstowych (stop lista, znaki interpunkcyjne, liczby itp.), sprowadzanie słów do postaci rdzenia znaczeniowego za pomocą algorytmu Portera oraz wybranych bibliotek informatycznych. Wyszukiwanie według słów kluczowych. Organizacja dokumentów w postaci macierzy term-dokument (ang. term-document matrix, TDM) oraz różne sposoby obliczania wagi poszczególnych termów (TF – term frequency, IDF – inverse document frequency). Miary podobieństwa wektorów oraz wykorzystanie ich do tworzenia rankingu wyszukanych dokumentów. Porównywanie jakości działania wyszukiwarek dokumentów tekstowych za pomocą różnych miar, np. precision-recall, krzywe ROC. Wybrane elementy algebry liniowej i zastosowanie ich do zadania aproksymacji macierzy TDM (ang. Low-rank approximation), omówienie korzyści z wykonanej aproksymacji. Różne techniki grupowania i klasyfikacji dokumentów. Ranking dokumentów oparty o strukturę połączeń: algorytm PageRank; autorytety i koncentratory. Tworzenie podsumowań dokumentów poprzez automatyczny wybór najważniejszych zdań oraz najważniejszych słów (termów). Tworzenie chmur słów (ang. wordclouds). Analiza wydźwięku dokumentów (np. pozytywny, negatywny, neutralny itp.). Modele wektorowe tekstów (ang. word embeddings). Systemy rekomendacyjne, (podejścia: item-based, user-based). Omówienie wybranych narzędzi informatycznych do realizacji zadań z dziedziny Text Mining.

Metody kształcenia

Wykład - wykład konwencjonalny.

Projekt - zajęcia praktyczne w laboratorium komputerowym.

Efekty uczenia się i metody weryfikacji osiągnięcia efektów uczenia się

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Potrafi zdefiniować pojęcie Text Mining oraz podać typowe przykłady zadań z tego obszaru wiedzy	• K_U01	• sprawdzian	• Wykład
Potrafi wykonywać wybrane analizy statystyczne dokumentów tekstowych	• K_U06	• przygotowanie projektu	• Projekt
Zna techniki wyszukiwania informacji tekstowych oraz tworzenia ich rankingu	• K_U16	• przygotowanie projektu	• Projekt
Potrafi stosować oprogramowanie eksploracji danych tekstowych	• K_U06 • K_U16	• przygotowanie projektu	• Projekt
Potrafi zdefiniować macierz TDM oraz zna techniki jej tworzenia i wykorzystywania w zadaniach wyszukiwania i rangowania dokumentów	• K_W02	• sprawdzian	• Wykład

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Zna bardziej zaawansowane techniki analizy danych tekstowych (systemy rekomendacyjne, analiza wydźwięku, modele wektorowe tekstów - word embeddings)	• K_W01 • K_W02	• sprawdzian	• Wykład

Warunki zaliczenia

Wykład – uzyskanie oceny pozytywnej ze sprawdzianu w formie pisemnej, przeprowadzonego na koniec semestru.

Projekt – ocena końcowa stanowi sumę ważoną ocen uzyskanych za realizację poszczególnych elementów projektu oraz formy jego prezentacji. Wkład poszczególnych elementów oceny: ocena projektu - 75%, wizualna forma prezentacji projektu 25%.

Ocena końcowa = 50 % oceny zaliczenia z formy zajęć wykład + 50 % oceny zaliczenia z formy zajęć projekt.

Literatura podstawowa

1. Markov, Z., Larose, D.T.: Eksploracja zasobów internetowych: Analiza struktury, zawartości i użytkowania sieci WWW, Warszawa, PWN, 2009.
2. Chakraborty, G., Pagolu, M., Garla, S.: Text Mining and Analysis: Practical Methods, Examples and Case Studies Using SAS, Cary, SAS Press, 2013.
3. Ch. Aggarwal, Ch-X Zhai, C. O'Neil: Mining Text Data, Springer, 2012.
4. D. Robinson, J. Silge: Text Mining with R, O'Reilly, 2017.
5. Julia Silge, David Robinson: Text Mining with R. A Tidy Approach, O'Reilly, 2017.
6. Michael W. Berry, Murray Browne: Understanding Search Engines: Mathematical Modeling and Text Retrieval, Society for Industrial and Applied Mathematics, 2005.
7. Dokumentacja systemu R (<https://www.r-project.org/>)

Literatura uzupełniająca

Uwagi

Zmodyfikowane przez dr hab. inż. Artur Gramacki, prof. UZ (ostatnia modyfikacja: 20-04-2022 23:48)

Wygenerowano automatycznie z systemu SylabUZ