Web mining - opis przedmiotu

Informacje ogólne

Nazwa przedmiotu	Web mining
Kod przedmiotu	11.3-WE-BizElP-EkspZasInter-Er
Wydział	Wydział Informatyki, Elektrotechniki i Automatyki
Kierunek	Biznes elektroniczny
Profil	praktyczny
Rodzaj studiów	Program Erasmus pierwszego stopnia
Semestr rozpoczęcia	semestr zimowy 2021/2022

Informacje o przedmiocie

Semestr	5
Liczba punktów ECTS do zdobycia	4
Typ przedmiotu	obowiązkowy
Język nauczania	angielski
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	-	-	Zaliczenie na ocenę
Projekt	30	2	-	-	Zaliczenie na ocenę

Cel przedmiotu

To familiarize students with basic models and techniques for discovering information found on the Internet
To familiarize students with text mining algorithms
Developing skills of exploring Internet resources based on statistical software.

Wymagania wstępne

Basics of statistics

Zakres tematyczny

Types of information on the internet. Introduction to Text Mining. Searching textual information. Preprocessing of text documents: removing unnecessary elements from text documents (stop list, punctuation, numbers, etc.), reducing words to the form of a semantic core using Porter's algorithm and selected IT libraries. Search by keywords. Organization of documents in the form of a term-document matrix (TDM) and various ways of calculating the weight of individual terms (TF - term frequency, IDF - inverse document frequency). Measures of similarity of vectors and using them to create a ranking of found documents. Comparing the quality of text document search engines using various measures, e.g. precision-recall, ROC curves. Selected elements of linear algebra and applying them to the task of TDM matrix approximation (Low-rank approximation), discussing the benefits of approximation. Various techniques for grouping and classifying documents. Document ranking based on connection structure: PageRank algorithm; authorities and hubs. Creating document summaries by automatically selecting the most important sentences and the most important words (terms). Creating wordclouds. Sentiment analysis as a technique to systematically identify, extract, quantify, and study affective states and subjective information (e.g. positive, negative, neutral, etc.). Presentation of selected IT tools for carrying out tasks in the field of Text Mining.

Metody kształcenia

Lecture, individual projects.

Efekty uczenia się i metody weryfikacji osiągania efektów uczenia się

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Is able to define the TDM matrix and knows the techniques of its creation and use in the tasks of searching and ranking documents.		• sprawdzian	• Wykład
Is able to perform selected statistical analyzes of text documents.		• przygotowanie projektu	• Projekt
Can use text data mining software.		• przygotowanie projektu	• Projekt
Is able to define the concept of Text Mining and give typical examples of tasks in this area of knowledge.		• sprawdzian	• Wykład
Knows the techniques of searching for text information and creating their ranking.		• przygotowanie projektu	• Projekt

Warunki zaliczenia

Lecture – the passing condition is to obtain a positive mark from the final test

Project– the passing condition is to obtain a positive mark from the project form

Calculation of the final grade: lecture 50% + project 50%

Literatura podstawowa

1. Larose, D.T.: Data Mining Methods and Models. John Wiley & Sons, New York 2006
2. Julia Silge, David Robinson:Text Mining with R. A Tidy Approach (available online: https://www.tidytextmining.com)
3. Markov, Z. and Larose, D.T.: Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage. John Wiley & Sons, New York 2007

Literatura uzupełniająca

Uwagi

Zmodyfikowane przez dr hab. inż. Marek Kowal, prof. UZ (ostatnia modyfikacja: 12-07-2021 11:41)

Wygenerowano automatycznie z systemu SylabUZ