

UČNI NAČRT PREDMETA / COURSE SYLLABUS

Predmet: Analiza tekstovnih in večpredstavnih podatkov ter semantične tehnologije
Course title: Text/Multimedia Mining and Semantic Technologies

Študijski program in stopnja Study programme and level	Modul Module	Letnik Academic year	Semester Semester
Informacijske in komunikacijske tehnologije, 3. stopnja	Tehnologije znanja	1	1
Information and Communication Technologies, 3 rd cycle	Knowledge Technologies	1	1

Vrsta predmeta / Course type

Izbirni / Elective

Univerzitetna koda predmeta / University course code:

IKT3-716

Predavanja Lectures	Seminar Seminar	Sem. vaje Tutorial	Lab. vaje Laboratory work	Druge oblike	Samost. delo Individ. work	ECTS
15	15			15	105	5

**Navedena porazdelitev ur velja, če je vpisanih vsaj 15 študentov. Drugače se obseg izvedbe kontaktnih ur sorazmerno zmanjša in prenese v samostojno delo. / This distribution of hours is valid if at least 15 students are enrolled. Otherwise the contact hours are linearly reduced and transferred to individual work.*

Nosilec predmeta / Lecturer:

Prof. dr. Dunja Mladenčić

Jeziki /

Predavanja / Lectures: Slovenščina, angleščina / Slovenian, English

Languages:

Vaje / Tutorial:

Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:

Zaključen študij druge stopnje s področja informacijskih ali komunikacijskih tehnologij ali zaključen študij druge stopnje na drugih področjih z znanjem osnov s področja predmeta. Potrebna so tudi osnovna znanja matematike, računalništva in informatike.

Prerequisites:

Completed second cycle studies in information or communication technologies or completed second cycle studies in other fields with knowledge of fundamentals in the field of this course. Basic knowledge of mathematics, computer science and informatics is also requested.

Vsebina:

Osnove dela s tekstovnimi in drugimi večpredstavnimi podatki:
 iskanje zakonitosti v podatkih
 osnovno procesiranje podatkov
 statistična dejstva in evidenca
 Predstavitev podatkov:
 leksikalna, sintaktična semantična
 Poglobljeni vpogled v izbrane scenarije zahtevnejše uporabe osnovnih metod:
 ekstrakcija znanja iz podatkov
 modeliranje uporabnikov

Content (Syllabus outline):

Basics of working with text and multimedia data:
 finding regularities in data
 data processing
 statistical artifacts vs. evidence
 Representation of data:
 lexical, syntactic, semantic
 Example tasks in complex data analytics:
 extracting information from text
 user modeling
 communication analysis
 multi-lingual and cross-lingual data

analiza komunikacije v omrežjih
večjezični in prekojezični podatki

Tehnike:
nadzorovano učenje
pol-nadzorovano učenje
ne nadzorovano učenje

Izzivi analize velikih količin podatkov:
netipične operacije nad podatki
shranjevanje velikih količin podatkov

Techniques:
supervised
semi-supervised
unsupervised learning

Handling data size:
atypical operators
storing big data

Temeljni literatura in viri / Readin

Izbrana poglavja iz naslednjih virov: / Selected chapters from the following sources:

- C. Sammut, and G. I. Webb, Eds. *Encyclopedia of Machine Learning*, Springer 2010. ISBN 0387307680 (selected entries)
- G. S. Ingersoll, T. S. Morton and A. L. Farris. *Taming Text: How to Find, Organize, and Manipulate It*, Manning Publications Co., 2013.
- J. Davies, M. Grobelnik, and D. Mladenić, Eds. *Semantic Web: Integrating Ontology Management, Knowledge Discovery and Human Language Technologies*, Springer, 2008. ISBN 3642100287 (selected chapters)
- P. Warren, J. Davies, and E. Simperl, Eds. *Context and Semantics for Knowledge Management: Technologies for Personal Productivity*. Heidelberg: Springer, cop. 2011. ISBN 3642195091 (selected chapters)
- C. C. Aggarwal. *Data Mining: The Textbook*, Springer, 2015 (selected chapters)

Cilji in kompetence:

Cilj predmeta je usposobiti študente za raziskovalno in razvojno delo na področju analize tekstovnih in večpredstavnih podatkov ter semantičnih tehnologij.

Kompetence študenta z uspešno zaključenim predmetom bodo vključevale razumevanje metod in tehnike analize tekstovnih in večpredstavnih podatkov ter semantičnih tehnologij in sposobnost za uporabo ustreznih orodij. Študent bo pridobil tudi znanja, potrebna za samostojno raziskovalno delo in razvoj.

Objectives and competences:

The goal of this course is to provide knowledge necessary for research and development in text and multimedia mining as well as in semantic technologies.

The competencies of the students completing this course successfully would include understanding of basic concepts, methods and techniques for text mining and cross modal data analysis as well as semantic technologies including usage of the relevant tools needed for research and development in the area.

Predvideni študijski rezultati:

Študenti bodo pridobili znanje in razvojne veščine na naslednjih področjih:

- predstavitev tekstovnih podatkov
- osnovni prijemi pri predprocesiranju tekstovnih in večpredstavnih podatkov
- metode za analizo tekstovnih in večpredstavnih podatkov
- vizualizacija podatkov
- osnovni pojmi iz področja semantičnih tehnologij

Intended learning outcomes:

The students will gain knowledge and skills in the following areas:

- data representation
- preprocessing text and multimedia
- text and multimedia mining methods
- data visualization
- basic concepts in semantic technologies
- ontologies and their usage in semantic technologies
- selected scenarios of using the presented

- izbrani scenariji uporabe predstavljenih metod
- osnove analize velikih količin podatkov

- methods
- basic issues in big data analytics

Metode poučevanja in učenja:

Predavanja, seminar, konzultacije, individualno delo

Learning and teaching methods:

Lectures, seminar, consultancy, individual work

Načini ocenjevanja:	Delež (v %) / Weight (in %)	Assessment:
Ustni izpit	50 %	Oral exam
Seminarska naloga	25 %	Seminar work
Ustni zagovor	25 %	Oral defense

Reference nosilca / Lecturer's references:

- **D. Mladenić**, and M. Grobelnik, Machine learning on text. In: GOLUB, Koraljka, Ed. Subject access to information : an interdisciplinary approach. Santa Barbara; Denver; Oxford: Libraries Unlimited, 2015, pp. 132-148.
- **D. Mladenić**, and M. Grobelnik, Automatic text analysis by artificial intelligence. *Informatica*, ISSN 0350-5596, 2013, 37:1, pp. 27-33.
- J. J. Starc, **D. Mladenić**. Joint learning of ontology and semantic parser from text. *Intelligent data analysis*, ISSN 1088-467X, 2017, vol. 21, no. 1, pp. 19-38.
- M. Karlovčec, B. Lužar, **D. Mladenić**. Core-periphery dynamics in collaboration networks : the case study of Slovenia. *Scientometrics*, ISSN 0138-9130, 2016, vol. 109, iss. 3, pp. 1561-1578.
- L. Bradeško, and **D. Mladenić**, A survey of chabot system through a Loebner prize competition. In: Erjavec, T. and Žganec Gros, J. (eds.) Proceedings of the Eighth Language Technologies Conference : proceedings of the 15th International Multiconference Information Society - IS 2012, volume C, Ljubljana: Institut Jožef Stefan, 2012, pp. 34-37.