

UČNI NAČRT PREDMETA / COURSE SYLLABUS

Predmet:	Analiza velikih količin podatkov
Course title:	Big Data Analysis

Študijski program in stopnja Study programme and level	Študijska smer Study field	Letnik Academic year	Semester Semester
Informatika v sodobni družbi, magistrski študijski program druge stopnje	-	Prvi ali drugi	Drugi ali četrty
Informatics in Contemporary Society, second cycle Masters Study Programme	-	First or second	Second or fourth

Vrsta predmeta / Course type

Izbirni / Elective

Univerzitetna koda predmeta / University course code:

1-ISD-MAG-IP-AVKP-2019-05-13

Predavanja Lectures	Seminar Seminar	Vaje Tutorial	Klinične vaje work	Druge oblike študija	Samost. delo Individ. work	ECTS
30	-	20	-	-	100	5

Nosilec predmeta / Lecturer:

Jeziki / Languages:

Predavanja / Lectures: Slovenski, angleški / Slovene, English

Vaje / Tutorial: Slovenski, angleški / Slovene, English

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

Ni formalnih predpogojev za pristop k predmetu; študentom se svetuje, da predhodno opravijo izpita iz Metod kvantitativne analize in/ali Podatkovnega rudarjenja.

Prerequisites:

No prerequisites are required, but students are strongly advised to complete the course of Quantitative analysis and/or Data mining prior to taking this course.

Vsebina:

- *Uvod v analizo velikih količin podatkov:* kaj so to veliki podatki, kje jih najdemo, kako jih shranimo?
- *Grafično predstavljanje velikih količin podatkov:* kateri diagrami so primerni za prikazovanje velikih količin podatkov;
- *MapReduce:* distribuirani datotečni sistemi in njihova organizacija, algoritmi, ki uporabljajo MapReduce;

Content (Syllabus outline):

- *Introduction to the big data analysis:* what is big data, where we find it, how to store it?
- *Visualizations of big data:* which diagrams are suitable for representing big data.
- *MapReduce:* Distributed File Systems; Large-Scale File-System Organization; Algorithms Using MapReduce; *Search for similar items:* near neighbour search,

- *Iskanje podobnih enot*: iskanje med najbližnjimi sosedi, povzemanje podatkov z ohranjanjem podobnosti, lokalno občutljive funkcije in razdalje;
- *Podatkovni tokovi*: podatkovni modeli za podatkovne tokove; vzorčenje podatkov; filtriranje podatkov v tokovih; štetje različnih enot v tokovih;
- *Analiza povezav*: PageRank algoritem; Nezaželenne povezave; Vozlišča in avtoritete;
- *Pogosti podatki*: štetje najpogostejših podatkov: ohranjanje pogostih podatkov v glavnem spominu;
- *Nadzorovano in nenadzorovano učenje*: razvrščanje v skupine, metoda podpornih vektorjev, regresijska analiza, klasifikacijska drevesa;
- *Uporaba visoko zmogljivih računalnikov za analizo velikih količin podatkov*.

- similarity preserving summaries of sets, locality sensitive functions and distances;
- *Data streams*: the stream data models; sampling data in a stream; filtering streams; counting distinct elements in a stream;
 - *Link analysis*: PageRank algorithm; Link spam; Hub and authorities;
 - *Frequent itemsets*: counting the frequent items in a stream, handling larger datasets in the main memory;
 - *Supervised and unsupervised learning*: clustering, support vector machines, regression analysis, classification trees;
 - *Use of high-performance computers to analyze big data*.

Temeljni literatura in viri / Readings:

- Leskovec, Jure, Rajaraman, Anand in Ullman, Jeffrey David (2014) *Mining of massive datasets*. New York: Cambridge university press.
- Trevor Hastie, Robert Tibshirani in Jerome Friedman (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Minelli, Michael, Chambers, Michele in Dhiraj, Ambiga (2013) *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*. Hoboken, New Jersey: John Wiley & Sons.
- Hiroshi Ishikawa (2015) *Social Big Data Mining*. CRC Press.

Cilji in kompetence:

Učna enota prispeva k razvoju naslednjih splošnih in predmetno specifičnih kompetenc:

- poznavanje in razumevanje kulturnih in družbenih procesov in sposobnost njihove kompleksne analize
- uporaba metodoloških orodij, tj. izvajanje, koordiniranje in organiziranje raziskav, uporaba raznih raziskovalnih metod in tehnik ter ocenitev njihove uporabnosti
- sposobnost sinteze izvornih idej, konceptov in rešitev določenih problemov iz različnih disciplinarnih področij

Objectives and competences:

The module contributes to the following general and subject-specific competences:

- familiarity with and understanding of cultural and social processes and competence for their complex analysis
- use of methodological tools, i.e. implementation, coordination and organisation of research, use of various research methods and techniques and to evaluate their usefulness
- competence to form original ideas, concepts and solutions for specific problems from different disciplines
- knowledge and understanding of a wide range of applications of

- poznavanje in razumevanje širokega nabora aplikacij informacijsko komunikacijske tehnologije v sodobni družbi
- poznavanje konceptov in metodologij za analizo velikih količino podatkov

- information communication technology in the modern society
- knowledge of the concepts and methodologies for the analysis of large amounts of data

Predvideni študijski rezultati:

- Znanje in razumevanje:
- Študent/študentka:*
- razume posebnosti analize velikih količin podatkov v primerjavi s klasičnimi podatkovnimi analizami
 - spozna metode, ki so primerne za analize tovrstnih podatkov in z uporabo zelo zmogljivih računalnikov in sodobne odprtokodne opreme izvede analizo na enem primeru velike količine podatkov

Intended learning outcomes:

- Knowledge and understanding:
- The student:*
- understands the specificity of big data analysis compared to classical data analysis
 - learns methods, designed for big data analysis and with the use of high performance computers and state of the art open source software analyse one instance of big data

Metode poučevanja in učenja:

- *predavanja* z aktivno udeležbo študentov (razlaga, diskusija, vprašanja, primeri, reševanje problemov, predstavitve)
- *vaje* v računalniški učilnici
- *individualne in skupinske konzultacije* (diskusija, dodatna razlaga, obravnava specifičnih vprašanj)

Learning and teaching methods:

- *lectures* (explanation with discussions, questions, case-studies, presentations)
- *tutorials* in the computer classroom
- *individual and group consultations* (debate, additional explanations, considering specific issues)

Načini ocenjevanja:

- Način (pisni izpit, ustno izpraševanje, naloge, projekt):
- samostojno pripravljena in predstavljena seminaraska naloga, v kateri študent naredi analizo enega vira velikih količin podatkov

Delež (v %) /
Weight (in %)

100

Assessment:

- Type (examination, oral, coursework, project):
- student individually prepares and presents a project related to analysis of one source of big data