

## UČNI NAČRT PREDMETA / COURSE SYLLABUS

<b>Predmet:</b>	Analiza velikih količin podatkov
<b>Course title:</b>	Big Data Analysis

Študijski program in stopnja Study programme and level	Študijska smer Study field	Letnik Academic year	Semester Semester
Računalništvo in spletne tehnologije, magistrski študijski program druge stopnje	-	Prvi	Drugi
Computer Science and Web Technologies, second cycle Masters Study Programme	-	First	Second

**Vrsta predmeta / Course type**

Izbirni / Elective

**Univerzitetna koda predmeta / University course code:**

2-RST-MAG-IP-AVKP-2019-03-05

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 izpit iz Napredne statistike in Diskretne matematike.

**Prerequisites:**

No prerequisites are required, but students are strongly advised to prior complete the courses of Advanced statistics and Discrete mathematics.

**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;
- sposobnost obvladovanja in pretvorbe realnega problema v obliki lažje predstavljivega modela;
- 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

#### Objectives and competences:

*The instructional unit contributes to the following general and subject-specific competences:*

- familiarity with and understanding of cultural and social processes and competence for their complex analysis;
- ability to manage and transform a real-life problem into the form of an easier to perceive model;
- use of methodological tools, i.e. implementation, coordination and organisation of research, use of various research methods and techniques and to evaluate their usefulness;

<p>problemov iz različnih disciplinarnih področij;</p> <ul style="list-style-type: none"> <li>• poznavanje in razumevanje širokega nabora aplikacij informacijsko komunikacijske tehnologije v sodobni družbi;</li> <li>• poznavanje konceptov in metodologij za analizo velikih količino podatkov.</li> </ul>	<ul style="list-style-type: none"> <li>• competence to form original ideas, concepts and solutions for specific problems from different disciplines;</li> <li>• knowledge and understanding of a wide range of applications of information communication technology in the modern society</li> <li>• knowledge of the concepts and methodologies for the analysis of large amounts of data.</li> </ul>
--	--

**Predvideni študijski rezultati:**

<p>Znanje in razumevanje:</p> <p><i>Študent/študentka:</i></p> <ul style="list-style-type: none"> <li>• razume posebnosti analize velikih količin podatkov v primerjavi s klasičnimi podatkovnimi analizami</li> <li>• 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</li> </ul>
---

**Intended learning outcomes:**

<p>Knowledge and understanding:</p> <p><i>The student:</i></p> <ul style="list-style-type: none"> <li>• understands the specificity of big data analysis compared to classical data analysis</li> <li>• 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</li> </ul>
---

**Metode poučevanja in učenja:**

<ul style="list-style-type: none"> <li>• <i>predavanja</i> z aktivno udeležbo študentov (razlaga, diskusija, vprašanja, primeri, reševanje problemov, predstavitve)</li> <li>• <i>vaje</i> v računalniški učilnici</li> <li>• <i>individualne in skupinske konzultacije</i> (diskusija, dodatna razlaga, obravnava specifičnih vprašanj)</li> </ul>
---

**Learning and teaching methods:**

<ul style="list-style-type: none"> <li>• <i>lectures</i> (explanation with discussions, questions, case-studies, presentations)</li> <li>• <i>tutorials</i> in the computer classroom</li> <li>• <i>individual and group consultations</i> (debate, additional explanations, considering specific issues)</li> </ul>
--

**Načini ocenjevanja:**

<p>Način (pisni izpit, ustno izpraševanje, naloge, projekt):</p> <ul style="list-style-type: none"> <li>• samostojno pripravljena in predstavljena seminarska naloga, v kateri študent naredi analizo enega vira velikih količin podatkov</li> </ul>
--

Delež (v %) /  
Weight (in %)

100 %

**Assessment:**

<p>Type (examination, oral, coursework, project):</p> <ul style="list-style-type: none"> <li>• student individually prepares and presents a project related to analysis of one source of big data</li> </ul>
--