

Студијски програми: Биоинформатика				
Врста и ниво студија: основне академске студије				
Назив предмета: Анализа великих база података				
Наставник (Име, средње слово, презиме): Данијела, Н.,Боберић Крстићев				
Статус предмета: изборни				
Број ЕСПБ: 7				
Услов: -				
<b>Циљ предмета</b>				
Циљ овог предмета је упознавање проблема присутних у анализи великих података и начина њиховог решавања користећи најновије доступне алате за обраду великих количина података у дистрибуираном окружењу.				
<b>Исход предмета</b>				
<i>Минимални:</i> На крају курса, се очекује да студенти схвате основне концепте и проблеме у анализи великих података и покажу способност да читају податке из постојећих извора података и учитају их у Apache Spark, као и да манипулишу тим подацима користећи програмски језик Python.				
<i>Пожељни:</i> Поред горепомнутих вештина и знања, успешан студент ће научити како да искажу алгоритме за анализу великих података у функционалном стилу како да избегну „shuffles“ и прерачунавање у Spark-у. Успешан студент ће такође разумети кључне проблеме и њихова решења као и њихову примену у практичном раду.				
<b>Садржај предмета</b>				
<i>Теоријска настава</i>				
Овај курс се фокусира на дистрибуирано програмирање користећи Spark и Python. На почетку курса се обрађује примена Python програмског језика и различитих библиотека за рад са великим подацима (Pandas, Matplotlib, Seaborn, Folium ...). Затим се дефинишу основни концепти и архитектура Spark фрејмворка. Детаљно се проучавају модули дефинисани у оквиру овог фрејмворка (SparkSQL, Spark Streaming, Machine learning with Spark). Такође, један део курса је посвећен обради података који се могу представити као граф. Кроз практичне примере у Spark-у, студенти уче када треба да се разматрају питања везана за дистрибуцију података и латенцију мрежне комуникације и како се у датом окружењу могу ефикасно решавати проблеми са добрим перформансама.				
<i>Вежбе</i>				
Практична настава се концентрише на коришћење уведених алатки за анализу великих података на већем броју примера и студија у решавању класичних проблема у пракси.				
<b>Литература</b>				
<i>Препоручена</i>				
<ol style="list-style-type: none"> <li>1. Andy Konwinski, Holden Karau, Matei Zaharia, and Patrick Wendell, “Learning Spark: Lightning-Fast Big Data Analysis”, O’Reilly, 2015.</li> <li>2. Balusamy, Balamurugan, Seifedine Kadry, and Amir H. Gandomi. Big Data: Concepts, Technology, and Architecture. John Wiley &amp; Sons, 2021.</li> <li>3. Viktor Mayer-Schönberger, Kenneth Cukier, “Big Data: A Revolution That Will Transform How We Live, Work, and Think”, Eamon Dolan/Mariner Books, 2013.</li> <li>4. Nandi, Amit. Spark for Python Developers. Packt Publishing Ltd, 2015.</li> <li>5. Guller, Mohammed. Big data analytics with Spark: A practitioner's guide to using Spark for large scale data analysis. Apress, 2015.</li> </ol>				
<b>Број часова активне наставе</b>				
Предавања:	Вежбе:	Практичне вежбе:	Студијски истраживачки рад:	Остали часови:
3	1	1	0	0
<b>Методe извођења наставе</b>				
На предавањима се користе класичне методе наставе уз коришћење пројектора и лаптопа, као и конекције на рачунарски кластер. Вежбе се састоје од већег броја примера који потичу из				

практичне употребе и где се ставља нагласак на самостално решавање проблема уз употребу најновијих технологија и алата као и рачунарског кластера.

**Оцена знања (максимални број поена 100)**

<b>Предиспитне обавезе</b>	<b>поена</b>	<b>Завршни испит</b>	<b>поена</b>
колоквијум	30	усмени испит	40
практични пројекат	30		