

Студијски програми: МАС Рачунарске науке			
Назив предмета: Анализа великих података			
Наставник: Данијела Боберић Крстићев			
Статус предмета: изборни			
Број ЕСПБ: 6			
Услов:			
Циљ предмета			
Циљ овог предмета је упознавање проблема присутних у анализи великих података и начина њиховог решавања.			
Исход предмета			
<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 горепоменутих вештина и знања, успешан студент ће научити како да примени различите алгоритме машинског учења као и алгоритме на графовима у циљу анализе великих података. Успешан студент ће такође разумети кључне проблеме и њихова решења као и њихову примену у практичном раду.			
Садржај предмета			
<i>Теоријска настава</i>			
На курсу ће бити речи о особинама и врстама великих података. Разматраће се методологија за анализу великих података. Курс ће се фокусирати на дистрибуирано програмирање користећи Apache Spark и програмски језик Python. Детаљно се приказује архитектура Spark развојног оквира као и појединачних модули који су део те архитектуре. Део курса биће посвећен визуализацији података користећи одговарајуће библиотеке у Python-у.			
<i>Практична настава</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. Кроз практичне примере, 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.			
Литература			
<ol style="list-style-type: none"> 1. Andy Konwinski, Holden Karau, Matei Zaharia, and Patrick Wendell, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly, 2015. 2. Nandi, Amit. Spark for Python Developers. Packt Publishing Ltd, 2015. 3. Guller, Mohammed. Big data analytics with Spark: A practitioner's guide to using Spark for large scale data analysis. Apress, 2015. 4. Zecevic, Petar, and Marko Bonaci. Spark in Action. Manning Publications Co., 2016. 5. Viktor Mayer-Schönberger, Kenneth Cukier, "Big Data: A Revolution That Will Transform How We Live, Work, and Think", Eamon Dolan/Mariner Books, 2013. 			
Број часова активне наставе	Теоријска настава: 2	Практична настава: 2	
Методe извођења наставе			
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. Знање студената се испитује на једном практичном колоквијуму и самосталном пројекту, док се на усменом делу испита проверава разумевање теоријског дела градива.			
Оцена знања (максимални број поена 100)			
Предиспитне обавезе	поена	Завршни испит	поена
Пројекат	30		
Колоквијум	30	усмени испит	40