

Студијски програм: Примењена математика – наука о подацима („Data Science“)				
Ниво студија: мастер студије				
Назив предмета: Анализа података у великим скалама				
Наставник: Душан Јаковетић, Милош Радовановић, Владимир Курбалија				
Статус: обавезни				
ЕСПБ: 5				
Услови: Препознавање облика и машинско учење, Теорија графова				
Циљ предмета				
<ul style="list-style-type: none"> - Увођење метода за рачунарску анализу података великих скала. - Савладавање програмирања и начина чувања података великих димензија као и њихова анализа. - Способност комбиновања вештина из области као што су чување података, дизајнирање дистрибуираних система, обрада сигнала, статистичка анализа података, машинско учење, теорија графова итд. у циљу формирања вредности које произилазе из великих података. 				
Исход предмета				
<ul style="list-style-type: none"> - Искуство у анализи и обради великих скупова података. - Способност дизајнирања и имплементације аналитичких решења: избор прикладног складиштења и алгорита, интерпретација и визуализација резултата. - Способност решавања проблема из разних области. 				
Садржај предмета				
<ul style="list-style-type: none"> - Складиштење података (Files, SQL, noSQL, Map-Reduce) и припрема; обрада упита; проналажење сличних предмета; анализа графова; Frequent itemset mining; селекција особина; Интеграција података/знања/метода (самостални, полусамостални и надгледани рад), приказ података. - Студије случаја и примене на хетерогене податке (текстови, социјални графови итд.) из реалних извора (smart phones, telecom operators, social media, satellite imagery, sensors, genomics) - Имплементација решења у програму Python са додатним пакетима: Numpy, SciPy, Networkx, Matplotlib, Orange, Scikit-learn, Pandas, PyMongo, Pydoop. 				
Литература				
<ol style="list-style-type: none"> 1. Jure Leskovic, Anand Rajaraman, Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2010. 2. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to data mining", Pearson Addison Wesley, 2006. 3. Jeffrey Dean, and Ghemawat Sanjay, "MapReduce: simplified data processing on large clusters", Communications of the ACM, 2008. 4. Santo Fortunato, "Community detection in graphs", Physics Reports, 2010. 5. Giovanni Seni, and John F. Elder, "Ensemble methods in data mining: improving accuracy through combining predictions", Synthesis Предавања on Data Mining and Knowledge Discovery, 2010. 6. Wes McKinney, Python for Data Analysis, O'Reilly Media, 2012. 				
Број часова активне наставе				Остало:
				0
Предавања: 2	Вежбе: 2	Остали облици наставе: 0	Студентски истраживачки рад: 0	
Методе извођења наставе				
Предавања; понављање; активно учешће студената у решавању проблема; тестови знања – колоквијум; домаћи; стварни подаци.				
Оцена знања (максимални број поена 100)				

Предиспитне обавезе		Поени	Завршни испит	Поени
домаћи	пројекат	60 = 30 (домаћи) + 30 (пројекат)	писмени испит	40