



# АКАДЕМИЈА СТРУКОВНИХ СТУДИЈА ЗАПАДНА СРБИЈА

Седиште Ужице, Трг Светог Саве 34

## О Д С Е К У ж и ц е

### План рада

Назив предмета	Истраживање података на Web-у				
Студијски програм/и (модул)	Мастер струковне студије Информационе технологије и системи				
Година студија	1	Семестар	2	ЕСПБ	7
Статус предмета	изборни		Услов	нема	

Подаци о наставницима и сарадницима на предмету	
Име и презиме, академско звање, термин консултација, електронска адреса	др Мирослава Јордовић Павловић, проф. стр. студија Консултације: четвртак, од 9 до 11 часова, кабинет 315 miroslava.jorodvic-pavlovic@vpts.edu.rs miroslavajp@gmail.com

Циљеви предмета
Циљ предмета је упознавање студената са основним техникама за рударење Web-а и информационих мрежа (укључујући социјалне мреже и социјалне медије). Конкретне могућности примене технологија Web mining-а.

Садржај и структура предмета
<p><i>Теоријска настава</i></p> <p>Увод у Web mining. Web crawling. Индексирање документа, рангирање. Екстракција информација и интеграција. Web mining алгоритми. Класификација и кластеризација докумената. Откривање образаца понашања корисника Web сајтова. Анализа социјалних мрежа. Алгоритми за истраживање социјалних мрежа. Анализа линкова. PageRank и SimRank алгоритми.</p> <p><i>Практична настава</i></p> <p>Упознавање са основним алатима Web mining-а и њихова примена. Алати за анализу линкова. Структура лог фајлова и алати за анализу логова.</p>

План и распоред извођења наставе	
Наставна недеља	НАЗИВ НАСТАВНЕ ЈЕДИНИЦЕ
<b>Предавања</b>	
1	Упознавање са обавезама на предмету. Увод у Web mining.
2	Web crawling.

3	Индексирање документа, рангирање.
4	Екстракција информација и интеграција.
5	Web mining алгоритми.
6	Web mining алгоритми.
7	Класификација и кластеризација докумената.
8	Класификација и кластеризација докумената.
9	Откривање образаца понашања корисника Web сајтова.
10	Анализа социјалних мрежа.
11	Алгоритми за истраживање социјалних мрежа.
12	Алгоритми за истраживање социјалних мрежа.
13	Анализа линкова.
14	PageRank и SimRank алгоритми.
15	PageRank и SimRank алгоритми.
<b>Вежбе</b>	
1	Упознавање са начином извођења вежби. Упознавање са опремом у рачунарској лабораторији. Библиотеке за web crawling. Визуализација колекције web страница као графа.
2	Практичан рад у рачунарској лабораторији: развој crawler-а за процесирање страница и издвајање образаца из сета страница.
3	Практичан рад у рачунарској лабораторији: екстракција информација. IR (Information Retrieval) модели.
4	Практичан рад у рачунарској лабораторији: приказ докумената, Weka Data Mining System
5	Практичан рад у рачунарској лабораторији: web mining алгоритми.
6	Практичан рад у рачунарској лабораторији: web mining алгоритми.
7	Практичан рад у рачунарској лабораторији: кластеризација докумената. K-means алгоритам.
8	Практичан рад у рачунарској лабораторији: класификација докумената. Nearest Neighbor алгоритам.
9	Практичан рад у рачунарској лабораторији: класификација докумената. Naive Bayes алгоритам.
10	Практичан рад у рачунарској лабораторији: алгоритми за истраживање социјалних мрежа.
11	Практичан рад у рачунарској лабораторији: алгоритми за истраживање социјалних мрежа.
12	Практичан рад у рачунарској лабораторији: алгоритми за истраживање социјалних мрежа.
13	Практичан рад у рачунарској лабораторији: анализа линкова
14	Практичан рад у рачунарској лабораторији: PageRank алгоритам.
15	Практичан рад у рачунарској лабораторији: SimRank алгоритам.

<b>Начин оцењивања – структура и број поена на предиспитним обавезама и испиту</b>			
<b>Предиспитне обавезе</b>	поена	<b>Завршни испит</b>	поена
активност у току предавања	5	писмени испит	25
практична настава	5	Пројекат	25
колоквијум-и	20	.....	
семинар-и	20		

## Литература

1. B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, Heidelberg; New York, 2011
2. M. Bonzanini, Mastering Social Media Mining with Python, Компјутер библиотека, 2016.
3. Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher Pal, Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems) 4th Edition, 2016.