

Факултет: ФАКУЛТЕТ ПО ТЕЛЕКОМУНИКАЦИИ И МЕНИДЖМЪНТ
Катедра: ИНФОРМАЦИОННИ ТЕХНОЛОГИИ
Професионално направление: 5.3. КОМУНИКАЦИОННА И КОМПЮТЪРНА ТЕХНИКА
Специалност: ИНФОРМАЦИОННИ ТЕХНОЛОГИИ
Образователно-квалификационна степен: МАГИСТЪР

ОПИСАНИЕ НА ЛЕКЦИОНЕН КУРС

1. Наименование на курса: *Извличане на знания от големи бази данни*

2. Код на курса: 23.2.4.0.22

3. Вид на курса: *избираем*

4. Ниво: *магистър*

5. Година на изучаване: *втора*

6. Семестър: *четвърти*

7. Брой кредити: 3

8. Име на лектора: *доц. д-р Красимира Иванова*

9. Резултати от обучението за дисциплината – усвоени знания, умения, компетенции (цели):

Студентите, приключили обучението си по дисциплината, трябва да придобият:

Основни знания за спецификата и използването на методите за извличане на знания от големи бази данни.

Практически умения за работа с алгоритми за извличане на знания от големи бази данни.

Целта на курса е да даде на студентите в систематизиран вид теоретични и практически знания за основните принципи и алгоритми за извличане на знания от големи бази данни.

10. Начин на преподаване: *лекции и практически упражнения*

11. Предварителни изисквания:

Необходими предварителни знания по дисциплините:

„Програмиране”, „Компютърни архитектури”, „Информационни технологии”, „Бази от данни в телекомуникациите”

12. Съдържание на курса (анотация):

Дисциплината съдържа: Извличане на знания – основни понятия и подходи. Ограничения. Методи за разпределено съхраняване и обработка на големи масиви данни. Разпределени файлови системи. Основни възможности на MapReduce. Откриване на подобни елементи. Търсене чрез метода на най-близкия съсед. Мерки за разстояние. Извличане на знания от потокови данни. Филтриране. Оценки. Анализ на взаимовръзки. Ранг на страници. Данни с често срещани се набори от елементи. Модел на пазарната кошница. Основни алгоритми за анализ. Работа с големи набори данни в основната памет. Основни методи и алгоритми. Въведение в методите за клъстеризация на големи обеми данни. „Проклятието” на размерността. Основни стратегии за клъстеризация. Йерархична клъстеризация в евклидови и неевклидови пространства. Ефективност.

13. Библиография (препоръчителна или задължителна литература)

Основна литература:

- Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman. Mining of Massive Datasets, Cambridge University Press, 2014. <http://www.mmnds.org/>

- Oded Maimon, Lior Rokach (Editors). Data Mining and Knowledge Discovery Handbook. Springer New York Dordrecht Heidelberg London, 2010. ISBN 978-0-387-09822-7 e-ISBN 978-0-387-09823-4; <http://romisatriawahono.net/lecture/dm/book/Maimon%20-%20Data%20Mining%20and%20Knowledge%20Discovery%20Handbook%20-%202010.pdf>

14. Методи и критерии на оценяване:

Дисциплината завършва с текуща оценка.

Проверяват се усвоените от студента знания по дисциплината и едновременно с това се оценяват неговите умения да мисли в реално време за решаване на поставени проблеми. Оценката се формира по точкова система.

Контролната работа по време на семестъра проверява продуктивните знания на студента, възможностите му да осмисля наученото.

Изисквания при формиране на оценката по шестобалната система

Отличен (6) – за добро познаване на информационните източници, задълбочено овладени ключови и допълнителни знания и умения, осмислено и правилно разбиране на материята, умения за решаване на сложни задачи, собствено мислене и аргументиране на решенията.

Мн. добър (5) – за много добре овладени ключови и допълнителни знания, осмислено и правилно разбиране на материята, умения за прилагане на наученото при сложни казуси и задачи.

Добър (4) – за овладени ключови и допълнителни знания за решаване на казуси и задачи, но без да може да ги развие до самостоятелно мислене.

Среден (3) – за усвоени ключови знания и решения на типови задачи.

Формиране на крайната оценка

	Компонент	Макс. точки	Пример 1	Пример 2
1	Контролна работа	30	20	16
2	Представяне на самостоятелно подготвен теоретичен казус	30	12	25
	Крайна оценка:	$0,1 * 60 = 6$	$\Sigma = 0,1 * 32 \approx 3$	$\Sigma = 0,1 * 41 \approx 4$

15. Език на преподаване: *български*