



Утвърдил:

Декан

Дата

СОФИЙСКИ УНИВЕРСИТЕТ "СВ. КЛИМЕНТ ОХРИДСКИ"

Факултет: Славянски филологии

Специалност: (код и наименование)

--	--	--	--	--	--	--	--	--

Магистърска програма: (код и наименование)

--	--	--	--	--	--	--	--	--

УЧЕБНА ПРОГРАМА

Дисциплина:

--	--	--	--

(код и наименование) Техники за анализ на данни за лингвисти

Преподавател: проф. д-р Мария Нишева

Асистент:

Учебна заетост	Форма	Хорариум
Аудиторна заетост	Лекции	30
	Семинарни упражнения	
	Практически упражнения (хоспетиране)	
Обща аудиторна заетост		30
Извънаудиторна заетост	Реферат	20
	Доклад/Презентация	
	Научно есе	
	Курсов учебен проект	20
	Учебна екскурзия	
	Самостоятелна работа в библиотека или с ресурси	20
Обща извънаудиторна заетост		60
ОБЩА ЗАЕТОСТ		90
Кредити аудиторна заетост		1
Кредити извънаудиторна заетост		2
ОБЩО ЕКСТ		3

№	Формиране на оценката по дисциплината ¹	% от оценката
1.	Workshops {информационно търсене и колективно обсъждане на доклади и реферати)	40
2.	Участие в тематични дискусии в часовете	
3.	Демонстрационни занятия	
4.	Посещения на обекти	
5.	Портфолио	
6.	Тестова проверка	
7.	Решаване на казуси	
8.	Курсов проект	60
9.		
10.		
11.		
12.	Изпит	

Анотация на учебната дисциплина:

Курсът запознава студентите с основните принципи и подходи от областта на анализа на данни и извличането на информация от текст. Въвеждат се основни понятия и методи за машинно самообучение, ориентирани към анализ и класификация на текст.

Студентите имат възможност да експериментират с имплементации на част от методите, разглеждани в курса, и да решават с тяхна помощ конкретни практически задачи.

Предварителни изисквания:

Необходимо е слушателите на курса да имат завършен успешно поне един граматически курс. Необходимо е и владеене на английски език на работно ниво.

Очаквани резултати:

Очаква се студентите, завършили успешно курса, да придобият базови знания и умения за анализ на текстови данни и извличане на информация от текст.

Учебно съдържание

№	Тема:	Хорариум
1.	Въведение в анализа на данни и извличането на информация. Основни понятия и задачи	2
2.	Разбиране на данни. Подготовка на данните за анализ	2
3.	Основни типове методи за анализ на данни. Представяне на несигурни данни с вероятности. Вероятности	4

¹ В зависимост от спецификата на учебната дисциплина и изискванията на преподавателя е възможно да се добавят необходимите форми, или да се премахнат ненужните.

	разсъждения. Статистически методи за анализ на данни	
4.	Предмет и цели на машинното самообучение. Класификация	2
5.	Бейсов класификатор. Класификация на текстове. Разпознаване на авторство	4
6.	Семантичен уеб и онтологии. Семантично търсене	4
7.	Извличане на мнения и анализ на настроения	2
8.	Анализ на използването на уеб страници	2
9.	Анализ на социални мрежи	2
10.	Софтуерни продукти за анализ на данни и извличане на информация. Orange	6

Конспект за изпит

№	Въпрос
1.	Изготвяне и защита на курсов проект, демонстриращ познания и умения в областта на анализа на текстови данни и извличането на информация от текст.

Библиография

Основна:

- Leskovec, J., A. Rajaraman, J. Ullman. Mining of Massive Datasets. Cambridge University Press, 2014.
- Witten, I., E. Frank, M. Hall, C. Pal. Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann, 2017.

Допълнителна:

- Hastie, T., R. Tibshirani, J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer, 2009.

Дата: 19.04.2024 г.

Съставил:
проф. д-р Мария Нишева