



УТВЪРДИЛ:

Декан

Дата

СОФИЙСКИ УНИВЕРСИТЕТ "СВ. КЛИМЕНТ ОХРИДСКИ"

Факултет: Славянски филологии

Специалност: (код и наименование)

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

.....
Магистърска програма: (код и наименование)

С	Л	Б	3	0					
---	---	---	---	---	--	--	--	--	--

Компютърна лингвистика. Интернет технологии в хуманитаристиката
(редовна форма на обучение)

УЧЕБНА ПРОГРАМА

Дисциплина:

--	--	--	--	--

Езикови бази от данни

Преподавател: доц. д-р Атанас Атанасов

Асистент:

Учебна заетост	Форма	Хорариум
Аудиторна заетост	Лекции	30
	Семинарни упражнения	
	Практически упражнения (хоспетиране)	
Обща аудиторна заетост		30
Извънаудиторна заетост	Реферат	
	Доклад/Презентация	
	Научно есе	
	Курсов учебен проект	30
	Учебна екскурзия	
	Самостоятелна работа в библиотека или с ресурси	30
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)	
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	50%
9.		
10.		
11.		
12.	Изпит	50%

Анотация на учебната дисциплина:

Целта на курса е да бъдат запознати студентите с основни положения при използването и създаването на бази от данни за лингвистични цели. Вниманието е насочено предимно към NoSQL бази данни, изградени върху XML технологии, като едно от най-мощните средства за кодиране и обработване на текстова информация.

Предварителни изисквания:

Базови знания за граматичната система на българския език; практически умения за анализ на основните граматични характеристики на текста; основни познания по XML.

Очаквани резултати:

Разширяване на теоретичните знания и придобиване на практически умения за работа с езикови бази от данни.

Учебно съдържание

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

№	Тема:	Хорариум
1.	Въведение в лингвистичните бази данни	2
2.	Общ преглед на SQL. Въведение в MySQL	2
3.	NoSQL бази данни, базирани на XML технологии	2
4.	XML бази от данни в eXist-DB	2
5.	Трансформация на езикови XML данни с XSLT	2
6.	Използване на шаблони в XSLT	2
7.	Цикли в XSLT	2
8.	Функции за обработване на текстови данни в XSLT	2
9.	Имплементация с отворен код на Xforms в eXist-DB: XSLTForms	2
10.	Регулярни изрази. Извличане и обработване на текстови низове	2
11.	Въведение в xQuery - заявки към XML бази от данни	2
12.	Създаване на езикови приложения в eXist-DB	2
13.	Индексиране на лингвистични бази от данни	2
14.	Структура на Принстънския WordNet и на Open Multilingual Wordnet	2
15.	Структура на данните в проекта FrameNet	2

Конспект за изпит

№	Въпрос
1	Изготвяне на курсов проект, демонстриращ познания в областта на езиковите бази от данни.

Библиография

Основна:

XPath Tutorial 2022: <http://www.w3schools.com/xpath/>

XSLT Tutorial 2022: <http://www.w3schools.com/xsl/>

Атанасов, А. 2016: SynTags - web interface for syntactic and semantic annotation. В: *Proceedings of the Second International Conference on Computational Linguistics in Bulgaria, Sofia, 2016*, 47-53, ISSN 2367-5578 (print), ISSN 2367-5675 (online)

Атанасов, А. 2016. Предложни фрази в предикативна позиция (аргументна структура) – *Littera et Lingua Series Dissertationes 10*, Издател: Факултет по

славянски филологии, Софийски университет „Св. Климент Охридски“, ISSN 1314-3352, <http://slav.uni-sofia.bg/naum/liliseriess/diss/2016/10>

Коева, С. 2010: Българският ФреймНет. София, 2010.

Коева, С., Влахова, Р., Декова, Р., Несторова, П., Атанасов, А. 2008: Българският ФреймНет. Семантико-синтактичен речник на българския език. София, 2008.

Рупенхофър 2006: Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson C.R. and Scheffczyk, J. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California.

Допълнителна:

Филмор 1982: Fillmore, C. J. Frame semantics. In: *Linguistics in the Morning Calm*, ed. by The Linguistic Society of Korea. Seoul: Hanshin, 1982, 111-137.

Дата: 25.02.2022 г.

Съставил:

доц. д-р Атанас Атанасов