



УТВЪРДИЛ:

Декан

Дата

СОФИЙСКИ УНИВЕРСИТЕТ "СВ. КЛИМЕНТ ОХРИДСКИ"

Факултет:

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:

--	--	--	--

Принципи на маркирането на съдържание

Преподавател: проф. д-р Андрей Бояджиев

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.	Текуща самостоятелна работа /контролно	
9.	Курсов проект	50%
10.		
11.		
12.	Изпит	50%

Анотация на учебната дисциплина:

Целта на курса е да бъдат запознати студентите с едно от най-мощните средства за обработване на текстови данни – езиците за маркиране, като бъде обърнато специално внимание на тяхното приложение в основните аспекти на компютърната лингвистика и хуманитаристика (електронни издания и описания, корпуси, търсещи машини, индексация и екстракция на данни и т.н.).

Предварителни изисквания:

Базови знания за строежа и структурата на текстовете, граматичната система на българския език; практически умения за анализ на основните граматични и структурни характеристики на текста.

Очаквани резултати:

Разширяване на теоретичните знания и практическите умения за работа с езиците за маркиране.

Учебно съдържание

№	Тема:	Хорариум
1	Същност на езиците за маркиране. Видове езици за маркиране	2
2	История и основи на XML	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.

3	Схеми и модели на XML	2
4.	Езиците XPath и XSLT	2
5.	Основни функции в XSLT	4
6.	HTML и XHTML	4
7.	Публикуване на HTML. Езикът CSS	4
8.	Уеб дизайн и бази данни	4
9.	XML и научните проекти в лингвистиката и хуманитаристиката	4

Конспект за изпит

№	Въпрос
1	Изготвяне на курсов проект, демонстриращ познания в областта на машинната обработка на съдържание от областта на текстовите корпуси.

Библиография

Основна:

- A Gentle Introduction to XML: <https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>
 CSS Tutorial: <http://www.w3schools.com/css/>
 HTML Tutorial: <http://www.w3schools.com/html/>
 XML Tutorial: <http://www.w3schools.com/xml/>
 XPath Tutorial: https://www.w3schools.com/xml/xpath_intro.asp
 XSLT Tutorial: https://www.w3schools.com/xml/xsl_intro.asp
 XQuery Tutorial: https://www.w3schools.com/xml/xquery_intro.asp
 Birnbaum, David. 2014. What is XML and why should humanists care? An even gentler introduction to XML: <http://dh.obdurodon.org/what-is-xml.xhtml>
 Бончев, Боян. 2015. *XML технологии*. София: Университетско издателство „Св. Климент Охридски“

Допълнителна:

- Алексова, Красимира и Антон Циммерлинг. 2020. Проект „Еволюция на граматическия строеж на българския и руския език в съпоставителен план: корпусен анализ и формални граматика“ <https://www.slav.uni-sofia.bg/evogramm/>
 Атанасов, Атанас. 2006. Проблеми при създаването на езикови корпуси с транскрибирана българска разговорна реч. В: *Паисиеви четения, Международен славистичен симпозиум*. Пловдив, 12 – 14.10.2006, 289-296.
 Атанасов, Атанас и Марина Джонова. 2014. Мултимедиен корпус на българската устна реч – структура и приложение, В: *Езикови ресурси и технологии за български език*, София, 2014, 127-141.
 Коева, Светла и Ивелина Стоянова. 2009. Български национален корпус. *Български език* 3. 137–146.
 Коева, Светла, Диана Благоева и Сия Колковска. 2011. Проектът Български национален корпус – резултати и перспективи. *Български език* 3. 34–53.

- Осенов, Петя и Кирил Симов. 2005. *Формална граматика на българския език*. София: БАН
<http://bultreebank.org/wp-content/uploads/2017/04/FormalGrammarBG.pdf>
- Търпоманова, Екатерина и Цветана Димитрова. 2014. Анотиране на паралелни многоезикови корпуси: Българско-английският паралелел корпус със съотнесени (прости) изречения (BulEnAC). В: *Езикови ресурси и технологии за български език*, 105 – 126. София: Академично издателство „Проф. Марин Дринов“.
- Bojadžiev, Andrej. 2012. Guidelines to Repertorium Initiative XML model for manuscript descriptions. *Scripta & e-Scripta* 10–11. 9 – 103.
- Bojadžiev, Andrej and Tsvetana Dimitrova. 2008. Linguistic Information in the Electronic Corpus of Old Slavic Texts. *Scripta & e-Scripta* 6. 105 – 151.

Дата:

Съставил:

проф. д-р Андрей Бояджиев