



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: 

--	--	--	--	--

Средства за създаване на езикови ресурси

**Преподавател:** доц. д-р Атанас Атанасов

Асистент:

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

№	Формиране на оценката по дисциплината <sup>1</sup>	% 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%
<b>Анотация на учебната дисциплина:</b>		
<p>Курсът има за цел да запознае студентите с някои от основните технологии и софтуерни продукти за създаване на електронни езикови ресурси.</p> <p>Акцентът пада върху практическото използване на XML като едно от най-мощните средства за кодиране на текстови данни.</p>		

<b>Предварителни изисквания:</b>
Базови знания за граматичната система на българския език; практически умения за анализ на основните граматични характеристики на текста.

<b>Очаквани резултати:</b>
Разширяване на теоретичните знания и практическите умения за създаване и използване на електронни езикови ресурси.

### *Учебно съдържание*

№	Тема:	Хорариум
1	Общ преглед на електронните езикови ресурси	2
2	Видове електронни речници	2
3	Грамматически речници	2

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

4.	Речниковият формат DELA	2
5.	Системата NooJ	2
6.	Създаване на речници в XML формат	6
7.	Стандартизиране на речници. Дефиниция на документен тип (DTD)	2
8.	Кодировки. Стандартът Unicode. Именувани символи (entities)	2
9.	EXMARaLDA - създаване на корпуси с разговорна реч	4
10.	Lexonomy - създаване и публикуване на речници	2
11.	Стандартът TEI	2
12.	Обработване на езикови XML данни с Python. Библиотеката lxml	2

### *Конспект за изпит*

№	Въпрос
1	Изготвяне на курсов проект, демонстриращ познания в областта на електронните езикови ресурси.

### *Библиография*

**Основна:**

- A Gentle Introduction to XML 2022: <http://www.tei-c.org/release/doc/tei-p4-doc/html/SG.html>
- Měchura, M. B. (2017): Introducing Lexonomy: an open-source dictionary writing and publishing system in *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 conference, 19-21 September 2017*, Leiden, The Netherlands.
- Атанасов, А. 2006: Проблеми при създаването на езикови корпуси с транскрибирана българска разговорна реч. В: *Паусиеви четения, Международен славистичен симпозиум*. Пловдив, 12 – 14.10.2006, 289-296.
- Атанасов, А. 2006: Създаване на учебни граматически електронни речници. В: *сб. доклади от Втората национална конференция с международно участие по електронно обучение във висшето образование*. Китен, 14-17.09.2006.
- Атанасов, Джонова 2014: Мултимедияен корпус на българската устна реч – структура и приложение, В: *Езикови ресурси и технологии за български език*, София, 2014, 127-141.
- Влахова, Р., Атанасов, А. 2009: Електронни езикови ресурси във Факултета по славянски филологии, СУ „Св. Климент Охридски, В: *сп. Български език*, кн. 3, 2009 г., 155-157.
- Коева, С. 2010: Българският семантично анотиран корпус. София, 2010.
- Паскалева, Е. 2007: Компютърна морфология. София, 2007.

***Допълнителна:***

Осенова, П., Симов, К. 2005: Моделиране на лингвистичните явления в синтактично аотиран корпус. В: *Аргументна структура. Проблеми на простото и сложното изречение*. СЕМАРШ, София, стр. 259-269.

Симов, К., Осенова, П., Видинска, П. 2008: Избор на словник и изработване на правописен речник на българския език на базата на корпуса VulTreeBank. В: *Изследвания по фразеология, лексикология и лексикография. В памет на проф. д-р Кети Анкова-Ничева*, стр. 407-414.

**Дата:** 25.02.2022 г.

**Съставил:**

доц. д-р Атанас Атанасов