



.....
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:

(код и наименование)

Езикови ресурси за езикови технологии

Преподавател: гл. ас. д-р Светлозара Лесева, гл. ас. д-р Мария Тодорова

Асистент:

Учебна заетост	Форма	Хорариум
Аудиторна заетост	Лекции	30
	Семинарни упражнения	
	Практически упражнения (хоспетиране)	
Обща аудиторна заетост		
Извънаудиторна заетост	Реферат	
	Доклад/Презентация	
	Научно есе	
	Курсов учебен проект	
	Учебна екскурзия	
	Самостоятелна работа в библиотека или с ресурси	60

Обща извънаудиторна заетост		
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	20%
3.	Демонстрационни занятия	
4.	Посещения на обекти	
5.	Портфолио	
6.	Тестова проверка	
7.	Решаване на казуси	
8.	Текуща самостоятелна работа / контролно	50%
9.		
10.		
11.		
12.	Изпит	30%

Анотация на учебната дисциплина:

Целта на обучението е студентите да се запознаят със съвременните езикови ресурси за български език и възможностите за тяхното приложение в компютърната обработка на естествения език и езиковите технологии. Курсът предлага цялостен подход към лингвистичното описание на ресурсите за български език и възможността за обогатяването им с нови видове анотация чрез съотнасянето на ресурси и надграждането им. Разглеждат се конкретни приложения на езиковите ресурси в сферата на езиковите технологии, обучението и изследователската работа.

Предварителни изисквания:

Няма.

Очаквани резултати:

Студентите ще добият представа за основни типове езикови ресурси с широко приложение в компютърната обработка на естествения език и

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

езиковите технологии. Ще се запознаят с различни нива на лингвистично описание и възможността за обогатяване на вече съществуващи ресурси с нови видове анотация с нови видове анотация чрез съотнасянето и надграждането им. Студентите ще имат възможност да се запознаят с конкретни приложения на езиковите ресурси в сферата на езиковите технологии.

Учебно съдържание

№	Тема:	Хорариум
1.	Езиковите ресурси в съвременната компютърна лингвистика	2
2.	Корпуси. Български национален корпус. Структура и метаописание. Общоезикови и специализирани корпуси. Едноезикови и многоезикови (паралелни и съпоставими) корпуси	2
3	Българският Браун корпус. Златен стандарт, тестов корпус, тренировъчен корпус	2
4	Паралелен българско-английски корпус, съотнесен на нивото на простите изречения в Българския национален корпус	2
5.	Лингвистична анотация – сегментация, морфологична анотация, синтактична анотация, семантична анотация; други нива на анотация (съставни единици, сложни форми, названия и имена и др.)	2
6.	Електронни речници – граматичен речник, морфологичен речник, речник на съставни лексикални единици	2
7.	Българският POS аотиран корпус. Дефиниране на тагово множество. Типове морфосинтактична многозначност	2
8.	Българският уърднет	2
9.	Българският семантично аотиран корпус	2
10.	Българският фреймнет	2
11.	Съотнасяне на ресурси за български език - Българският фреймнет и Българският уърднет	2
12.	Обогатяване на езиковите ресурси с допълнителни нива на описание - съставни лексикални единици, наименования, термини, сложни глаголни форми,	2

	концептуална информация (концептуални фреймове)	
13.	Области на приложение на езиковите ресурси в сферата на езиковите технологии	2
14.	Области на приложение на езиковите ресурси в сферата на образованието	2
15.	Области на приложение на езиковите ресурси в сферата на изследователската работа	2

Конспект за изпит

№	Въпрос
1.	Курсова работа
2.	Представяне на курсовата работа и събеседване

Библиография

Основна:

Коева, С. Граматичен речник на българския език. Описание на концепцията за организацията на лингвистичните данни. – *Български език*, 1998, (6), 49-58.

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

Коева, С. (съст.). *Езикови ресурси и технологии за български език*. Академично издателство „Проф. Марин Дринов“, 2014.

Коева, С. Изследователски подходи в науката за езика и езиковото обучение. – *Български език и литература*, LIX, 5, 2017, ISSN:0323 – 9519, 518-533

Коева, С., С. Лесева, И. Стоянова, М. Тодорова. Езиковите технологии и ресурси – нови перспективи в обучението по български език (Българската лексикално-семантична мрежа Булнет и Българският национален корпус). – *Български език и литература*, кн.4, 2016, с. 377 – 393. ISSN:0323–9519.

Стоянова, И., С. Лесева. Автоматично разпознаване на прости изречения в състава на сложното в българския език. – В: С. Коева, П. Бъркалова (съст.). *Доклади от Юбилейната научна сесия „Съвременни тенденции в езиковедските изследвания“*, посветена на 85 години от рождението на проф. д.ф.н. Йордан Пенчев, 10 – 11 ноември, 2016 г., Издателство на БАН „Проф. Марин Дринов“, 2017, ISBN:978-954-322-917-8, 360-375.

Тодорова, М., С. Лесева, И. Стоянова. Речник на съставните думи в българския език – развитие и перспективи. – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“ (София, 15 – 16 май 2017 година)*, 1, Институт за български език „Проф. Любомир Андрейчин“, 2017, ISBN:978-954-924-899-9, 311-319.

Тодорова, М., С. Лесева. БНК – области на приложение. – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“ (София, 15 – 16 май 2017 година)*, 2, Институт за български език „Проф. Любомир Андрейчин“, 2017, ISBN:978-954-924-899-9, 243-252.

Koeva, S., I. Stoyanova, S. Leseva, T. Dimitrova, R. Dekova, and E. Tarpomanova. The Bulgarian National Corpus. Theory and Practice in Corpus Design. – In: *Journal of Language Modeling*, vol. 0, Nr. 1, 2012, 65-110. <http://jlm.ipipan.waw.pl/index.php/JLM/article/view/33/30>

Koeva, S., I. Stoyanova, M. Todorova, S. Leseva, T. Dimitrova. Metadata Extraction, Representation and Management within the Bulgarian National Corpus. – In: *Proceedings of the 4th Workshop on Challenges in the Management of Large Corpora Workshop Programme*, ELDA, 2016, 33-39.

Koeva, S., I. Stoyanova, M. Todorova, S. Leseva. Semi-automatic Compilation of the Dictionary of Bulgarian Multiword Expressions. – *Proceedings of the GLOBALEX 2016 Workshop: Lexicographic Resources for Human Language Technology*, LREC, 2016, 86-95.

Koeva, S. Online Language Resources and Technology Tools provided by the Institute for Bulgarian Language. Scientific Works of University of Food Technologies. – In: *Proceedings of the 65rd Scientific Conference with International Participation "Food Science, Engineering and Technology – 2018"*, 65, 1, Plovdiv: University of Food Technologies, 2019, ISSN:1314-7102 CD version 2535-1311 online version, 272-284.

Koeva, S., D. Hristov, T. Dimitrova, V. Stefanova. Enriching Wordnet with Frame Semantics. – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“ (София, 14 – 15 май 2019 година)*. Ваня Мичева, Диана Благоева, Сия Колковска, Татяна Александрова, Христина Дейкова (отг. ред.), София: Издателство на БАН „Проф. Марин Дринов“, 2019, ISBN:978-954-322-987-1, 300-308.

Koeva, S., T. Dimitrova, D. Hristov, H. Kukova, S. Leseva, V. Stefanova, I. Stoyanova, M. Todorova. *Towards a Semantic Network Enriched with a Variety of Semantic Relations*. Sofia: Prof. Marin Drinov Academic Publishing House of the Bulgarian Academy of Sciences, 2020.

Koeva, S. Bulgarian wordnet – How does it differ from the Princeton wordnet. – *Papers of BAS*, 8, 1, Prof. Marin Drinov Publishing House of Bulgarian Academy of Sciences, 2021, ISSN:2367-6248, 47-70.

Leseva, S., I. Stoyanova, M. Todorova, H. Kukova. A Theoretical Overview of Conceptual Frames and Semantic Restrictions on Frame Elements. – *Балканско езиковедие/Linguistique Balkanique*, LVIII, 2, 2019, ISSN:0324-1653, 172-186

Leseva, S., Stoyanova, I.. Enhancing Conceptual Description through Resource Linking and Exploration of Semantic Relations. – In: *Proceedings of the Tenth Global Wordnet Conference* (July 23–27, 2019, Wrocław (Poland), Oficyna Wydawnicza Politechniki Wrocławskiej, 2019, ISBN:978-83-7493-108-3, 280-289

Допълнителна:

Коева, С., В. Стефанова. Меронимията в УърдНет: дефиниране на субрелации. – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“* (София, 2020). Ваня Мичева, Диана Благоева, Марияна Витанова, Марияна Цибранска, Сия Колковска, Татяна Александрова (отг. ред.), 2, Издателство на БАН „Проф. „Марин Дринов“, 2020, 212-223

Кукова, Х. Глаголи за комуникация, фреймови елементи и семантични ограничения (върху синонимни множества от Булнет). – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“* (София, 2020). Ваня Мичева, Диана Благоева, Марияна Витанова, Марияна Цибранска, Сия Колковска, Татяна Александрова (отг. ред.), 2, Издателство на БАН „Проф. Марин Дринов“, 2020, ISSN:2683-118X (print); 2683-1198 (online), 233-241

Лесева, Св., Ив. Стоянова, Хр. Кукова, М. Тодорова. Интегриране на субкатегоризационна информация в релационната структура на УърдНет. – *Български език*, 65, 2, 2018, ISSN:0005-4283, 11-40

Стефанова, В. Класификация на прилагателните имена в VulNet. – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“* (София, 15 –16 май 2017 г.), 2, Институт за български език „Проф. Любомир Андрейчин“, 2017, ISBN:978-954-92489-9-9, 201-207

Koeva, S., T. Dimitrova, V. Stefanova, D. Hristov. Mapping WordNet concepts with CPA ontology. – In: *Proceedings of the 9th Global WordNet Conference (GWC'2018)*, Global Wordnet Association, Singapore, 2018, ISBN:978-981-11-7087-4, 70-77. SJR (Scopus):0.19

Leseva, S., I. Stoyanova, M. Todorova. Towards the Automatic Identification of Light Verb Constructions in Bulgarian. *Proceedings of the Second International Conference Computational Linguistics in Bulgaria (CLIB 2016)*, The Institute for Bulgarian Language Prof. Lyubomir Andreychin, Bulgarian Academy of Sciences, 2016, 28-37

Светлозара Лесева, Мария Тодорова