



Утвърдил:

Декан

Дата

СОФИЙСКИ УНИВЕРСИТЕТ “СВ. КЛИМЕНТ ОХРИДСКИ”

Факултет: Факултет по славянски филологии

Специалност: (код и наименование)

| | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|

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

| | | | | | | | | | |
|---|---|---|---|---|--|--|--|--|--|
| С | Л | Б | З | 0 | | | | | |
|---|---|---|---|---|--|--|--|--|--|

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

УЧЕБНА ПРОГРАМА

Дисциплина:

| | | | | |
|--|--|--|--|--|
| | | | | |
|--|--|--|--|--|

Корпусна лингвистика: компютърни подходи

(задочно обучение)

Преподавател: гл. ас. д-р Цветана Димитрова

| Учебна заетост | Форма | Хорариум |
|------------------------------------|--------------------------------|-----------|
| Аудиторна заетост | Лекции | 20 |
| Обща аудиторна заетост | | 20 |
| | Учебен проект | 35 |
| | Самостоятелна работа с ресурси | 35 |
| Обща извънаудиторна заетост | | 70 |
| ОБЩА ЗАЕТОСТ | | 90 |
| Кредити аудиторна заетост | | 1 |
| Кредити извънаудиторна заетост | | 2 |
| ОБЩО ЕКСТ | | 3 |

| № | Формиране на оценката по дисциплината ¹ | % от оценката |
|---|--|---------------|
| | Самостоятелна работа с ресурси | 40 |
| | Изпит | 60 |

Анотация на учебната дисциплина:

Предлаганият курс запознава студентите с методите на корпусната лингвистика, която осигурява данни за създаването на езикови технологии. Представят се основните понятия,

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

методи и технологии за създаване и работа с езикови корпуси, съществуващите корпуси (едноезикови и многоезикови) за български и за други езици, както и с основните и технологии за създаване и обработка на корпусни данни и за търсене в корпуси. Целта на курса е студентите да придобият умения за работа с корпуси: да могат да създават корпуси; да могат да използват информацията от метаописанието и нивата на анотация (лингвистична и екстралингвистична) на корпусите; да се запознаят с принципите на синтаксиса и семантиката на заявките, които се използват за търсене в корпусни данни. Разглеждат се и приложенията на корпусите в различни езиковедски и други направления.

Предварителни изисквания:

Няма.

Очаквани резултати:

Студентите ще придобият теоретични знания и практически умения за работа с различни корпуси, както и за съставяне на корпуси - подбор и обработка на подходящ езиков материал (включително за конкретни задачи) и трансформацията му в данни. Ще се запознаят с основните функционалности на различни корпусни ресурси и с програми и приложения за работа с тях, включително за извършване на справки и за извличане на различна информация от тях.

Учебно съдържание

| № | Тема: | Хорариум |
|----|---|----------|
| 1 | Езикови корпуси: същност, история, основни понятия | 2 |
| 2 | Видове езикови корпуси | 2 |
| 3 | Съставяне на езикови корпуси | 2 |
| 4. | Корпусни метаданни (видове, стандарти) | 2 |
| 5 | Анотация на корпуси (видове, стандарти) | 2 |
| 6 | Българският национален корпус (БНК); подкорпуси в състава на БНК | 2 |
| 7 | Многоезикови корпуси; съпоставими и паралелни корпуси | 2 |
| 8 | Търсене в корпусни данни (регулярни изрази; заявки; приложения) | 2 |
| 9 | Статистика и корпусни данни | 2 |
| 10 | Съставяне и обработка на езикови корпуси – практическа работа: приложения (SketchEngine, AntCorGen./AntConc/AntGram и др.), подходи; приложение за изследователски цели | 2 |

Конспект за изпит

| № | Въпрос |
|---|--------------------------------------|
| 1 | Разработване на курсов учебен проект |
| 2 | Умения за работа с корпуси |

Библиография

Основна:

- A Practical Handbook of Corpus Linguistics. Edited by Magali Paquot, Stefan Th. Gries. Springer, 2020.
- The Routledge Handbook of Corpus Linguistics. Edited by Anne O'Keeffe, Michael J. McCarthy. 2nd edition. 2022.
- Weisser, Martin. Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis. Wiley-Blackwell, 2015.
- Developing Linguistic Corpora: a Guide to Good Practice. Edited by Martin Wynne. Oxford: Oxbow Books (2005) <<https://users.ox.ac.uk/~martinw/dlc/index.htm>>
- McEnery, Tony, Andrew Wilson. Corpus Linguistics. Edinburgh: Edinburgh University Press, 2001.
- Sinclair, John. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.
- Kennedy, Graeme. An Introduction to Corpus Linguistics. London: Routledge, 1998 (e-book 2014).
- Corpus Linguistics. An International Handbook. Volumes 1-2. Edited by Anke Lüdeling, Anke, Merja Kytö. Berlin & New York: Walter de Gruyter, 2008.
- Leech, Geoffrey. Corpus Annotation Schemes. // Literary and Linguistics Computing. 1993. Vol. 8. N 4, 275-281.
- Kilgarriff, Adam, Gregory Grefenstette. Web as Corpus. 2003. <<http://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf>>
- Brezina, Vaclav. Statistics in Corpus Linguistics: A Practical Guide. Cambridge: Cambridge University Press, 2018. <<http://corpora.lancs.ac.uk/stats/index.php>>
- Лесева, Светлозара, Мария Тодорова. Българският национален корпус – области на приложение. // Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин“ (София, 15 – 16 май 2017 година). Част 1. <http://ibl.bas.bg/Proceedings_IBL_Conference_2017_Volume_I.pdf>
- Коева, Светла. Езикови ресурси и технологии за български език. София: Акад. изд. „Проф. Марин Дринов“, 2014.
- Koeva, Svetla, Ivelina Stoyanova, Svetlozara Leseva, Tsvetana Dimitrova, Rositsa Dekova, Ekaterina Tarpomanova. The Bulgarian National Corpus: Theory and Practice in Corpus Design. Journal of Language Modelling, 1, 2012, 65-110. <<http://dx.doi.org/10.15398/jlm.v0i1.33>>
- Коева, Светла. Лингвистична анотация. Известия на Института за български език „Проф. Любомир Андрейчин“, 2014, 7-34. <<https://ibl.bas.bg/izvestiya/wp-content/uploads/2018/06/IZVESTIA-IBE-XXVII-7-33-Koeva.pdf>>

Допълнителна:

- BNC: The BNC Users Reference Guide, 2000. <<http://www.natcorp.ox.ac.uk/World/HTML/>>
- EAGLES Documentation 1996. <<http://www.ilc.cnr.it/EAGLES/edintro/edintro.html>>
- Corpus building and investigation for the Humanities: An on-line information pack about corpus investigation techniques for the Humanities.

<https://www.birmingham.ac.uk/research/activity/corpus/publications/introduction-corpus-investigative-techniques.aspx>

Лесева, Светлозара, Ивелина Стоянова, Мария Тодорова, Светла Коева. Езиковите технологии и ресурси – нови перспективи в обучението по български език (Българската лексикалносемантична мрежа Булнет и Българският национален корпус). – Български език и литература, кн. 4, 2016, с. 377 – 393.

Дата: 2022 г.

Съставил:

Цветана Димитрова