



ЦЕНТЪР ЗА ОБУЧЕНИЕ – БАН

1000 София
ул. „Сердика“ № 4
<http://edu.bas.bg>

email: tdc-phd@cu.bas.bg
тел.: 02 987 31 67
02 979 52 60

Основна информация:

Име на курса: Увод в компютърната лингвистика

Лектор: проф. д-р Светла Коева

Телефон: +35929722302

Имейл: svetla@dcl.bas.bg

Хорариум: 30 часа

Анотация (до 150 думи):

Лекционният курс запознава докторантите с основните алгоритми в съвременната компютърна лингвистика и с някои от най-важните им приложения. Учебният материал ще бъде представен достъпно, така че за неговото овладяване да не се изисква предварителна подготовка по математика и/или информатика. Докторантите ще се запознаят с основите на компютърната лингвистика като теоретична и приложна област и ще получат познания за нейните задачи, методи и резултати, за да може да участват успешно в по-задълбочени разработки в рамките на бъдещата си професионална дейност.

Тематично съдържание на курса (кратко описание по теми или модули):

#	Тема	Часове
1	Компютърна лингвистика: обхват, задачи и приложения	2
2	Теория на формалните езици: йерархията на Чомски	2
3	Регулярни изрази и крайни автомати	2
4	Контекстносвободни и контекстнозависими граматика	2
5	Нива на езикова обработка: токънизация, лематизация, синтактичен анализ и автоматично отстраняване на многозначност	2
6	Векторна семантика и вграждане на думи	2
7	Архитектури на невронни мрежи за компютърна обработка на езика	2
8	Трансформатори и големи езикови модели	2
9	Диалогови системи и разговорни агенти	2
10	Проектиране на инструкции за големи езикови модели	2
11	Машинен превод и компютърно подпомогнат превод	2
12	Автоматичен анализ и синтез на реч	2
13	Лингвистична анотация: принципи и практика	2



ЦЕНТЪР ЗА ОБУЧЕНИЕ – БАН

1000 София
ул. „Сердика“ № 4
<http://edu.bas.bg>

email: tdc-phd@cu.bas.bg
тел.: 02 987 31 67
02 979 52 60

14	Автоматично свързване на анафори и антецеденти	2
15	Семантично моделиране и представяне на знание	2
Общо		30

Форми на обучение и оценяване:

Курсът включва лекции и практически упражнения. Лекциите представят теоретичните основи и ключовите понятия на всяка тема, а упражненията дават възможност на докторантите да прилагат инструменти и методи за обработка на езикови данни, като затвърждават знанията си чрез непосредствена работа с материала.

Оценяването се осъществява чрез курсова работа. Докторантите разработват индивидуален проект, в който прилагат един или няколко от разгледаните в курса методи за решаване на самостоятелно избран проблем, свързан по възможност с тяхното дисертационно изследване. Проектът се придружава от писмена курсова работа, в която докторантът представя използваните методи, начина на тяхното прилагане и критичен анализ на получените резултати.

Компетентности, придобити в резултат на обучението (3-5 точки):

След успешно завършване на курса студентите ще могат да:

Демонстрират систематично разбиране на основните задачи, методи и теоретични основи на компютърната лингвистика, включително теорията на формалните езици, приложенията за обработка на текст и съвременните невронни мрежи.

Проектират и изпълняват малък проект по компютърна лингвистика – от формулиране на проблема и избор на метод до прилагане, анализ и писмено представяне на резултатите.

Работят с научна литература в областта на компютърната лингвистика и сродните дисциплини, като ситуират докторските си изследвания в по-широкия контекст на езиковите технологии и лингвистичната теория.

Разсъждават критично върху възможностите и ограниченията на съвременните езикови технологии, включително въпроси, свързани с качеството на данните, пристрастността на моделите и връзката между компютърните модели и лингвистичната теория.

Литература:

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, 33. Curran Associates, Red Hook, NY, 2020, pp. 1877–1901.

Carta, S., Giuliani, A., Piano, L., Podda, A. S., Pompianu, L., & Tiddia, S. G. Knowledge graph construction: Extraction, learning, and evaluation. *Applied Sciences*, 15(7), Article 3727. MDPI, Basel, 2025.

de Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. Universal Dependencies. *Computational Linguistics*, 47(2) MIT Press, Cambridge, MA, 2021, pp. 255–308.

Jurafsky, Daniel and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing*, 3rd edition. Online manuscript released January 12, 2025.

Lane, Hobson and Maria Dyshel. 2025. *Natural Language Processing in Action*. 2nd edition. MEAP.



ЦЕНТЪР ЗА ОБУЧЕНИЕ – БАН

1000 София
ул. „Сердика“ № 4
<http://edu.bas.bg>

email: tdc-phd@cu.bas.bg
тел.: 02 987 31 67
02 979 52 60

Manning, C. D., and Schutze, H. 1999. *Foundations of statistical natural language processing*. MIT Press.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*. PMLR, Honolulu, 2023, pp. 28492–28518.

Rothwell, A., Way, A., & Youdale, R. *Computer-Assisted Literary Translation*. Routledge, London, 2023.

Russell, Stuart J. and Peter Norvig. 2022. *Artificial Intelligence: A Modern Approach*, 4th edition. Pearson.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*, 30. Curran Associates, Red Hook, NY, 2017, pp. 5998–6008.

Коева, Св., Д. Благоева (ред.). *Езикови ресурси и технологии за български език*. София: Академично издателство „Проф. “Марин Дринов”, 2014, 310 с. ISBN: 978-954-322-797-6.

Допълнителна информация (по желание) (например: специални изисквания, лабораторно оборудване, предварителни знания):

Няма.

Курсът е предназначен предимно за езиковеди, но може да бъде посещаван от специалисти от други хуманитарни дисциплини, математици и информатици., които искат да разширят своята перспективи.