

АННОТАЦИЯ
дисциплины Б1.В.04 «Основы компьютерной лингвистики»

Направление подготовки/специальности

02.03.02 Фундаментальная информатика и информационные технологии.

Объём трудоёмкости: 2 зачетных единицы (72 часа, из них – 34 часа аудиторной нагрузки: лекционных 16 ч, лабораторных 18 ч, КСР 2 ч, ИКР 0,2 ч. СРС 35,8 ч., включая 3,8 ч подготовку к контрольным мероприятиям).

Цель дисциплины:

дать студентам базовые знания по основным подходам к построению моделей и автоматической обработке естественных языков, а также приложениям в обработке текстовой информации на естественном языке, научить решать комплексные задачи в области проектирования систем обработки нечисловой информации.

Задачи дисциплины:

В результате освоения дисциплины должны быть решены следующие основные задачи. Студент должен:

- знать базовые сведения по основным подходам к построению моделей и автоматической обработке естественных языков, а также приложениям в обработке текстовой информации на естественном языке, приобрести навыки решения комплексных задач в области проектирования систем обработки нечисловой информации.

- уметь применять знания по основным подходам к построению моделей и автоматической обработке естественных языков, а также приложениям в обработке текстовой информации на естественном языке в области проектирования систем обработки нечисловой информации.

- владеть восприятием, анализом и обобщением информации в профессиональной области и выбором путей решения профессиональных задач на основе знаний и умений дисциплины «Основы компьютерной лингвистики».

Место дисциплины в структуре ООП ВО

Курс «Основы компьютерной лингвистики» относится к вариативной части блока Б1 учебного плана, формируемой участниками образовательных отношений. Для изучения дисциплины необходимо знание языков и методов программирования, дискретной математики, также базовых сведений по грамматике русского языка. Знания, получаемые при изучении курса, используются при изучении таких дисциплин цикла учебного плана бакалавра, как «Модели интеллектуальных систем», «Методы разработки трансляторов».

Требования к уровню освоения дисциплины

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора*	Результаты обучения по дисциплине (знает, умеет, владеет (навыки и/или опыт деятельности))
УК-4 Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(-ых) языке(-ах)	
УК-4.1. Знает литературную форму государственного языка, основы устной и письменной коммуникации на иностранном языке.	Знает литературную форму государственного языка, основы устной и письменной коммуникации на иностранном языке в области компьютерной лингвистики
УК-4.2. Умеет выражать свои мысли на государственном, родном и иностранном языке в ситуации деловой коммуникации.	Умеет выражать свои мысли на государственном, родном и иностранном языке в ситуации деловой коммуникации в области компьютерной лингвистики.
УК-4.3. Имеет практический опыт составления текстов разной функциональной принадлежности и разных жанров на государственном и родном языках, опыт перевода текстов с иностранного языка на родной, опыт говорения на государственном и иностранном языках.	Имеет практический опыт составления текстов разной функциональной принадлежности и разных жанров на государственном и родном языках, опыт перевода текстов с иностранного языка на родной, опыт говорения на государственном и иностранном языках в области компьютерной лингвистики.

Код и наименование индикатора*	Результаты обучения по дисциплине (знает, умеет, владеет (навыки и/или опыт деятельности))
ПК-2 Способен проводить под научным руководством локальные исследования на основе существующих методов в конкретной области профессиональной деятельности	
ПК-2.1. Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации владеет навыками подготовки научных обзоров, публикаций, рефератов и библиографий по тематике проводимых исследований на русском и английском языке.	Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации владеет навыками подготовки научных обзоров, публикаций, рефератов и библиографий по тематике проводимых исследований на русском и английском языке в области компьютерной лингвистики.
ПК-2.2. Умеет решать научные задачи в связи с поставленной целью и в соответствии с выбранной методикой.	Умеет решать научные задачи в связи с поставленной целью и в соответствии с выбранной методикой в области компьютерной лингвистики.
ПК-2.3. Имеет практический опыт выступлений и научной аргументации при анализе объекта научной и профессиональной деятельности.	Имеет практический опыт выступлений и научной аргументации при анализе объекта научной и профессиональной деятельности в области компьютерной лингвистики.

Результаты обучения по дисциплине достигаются в рамках осуществления всех видов контактной и самостоятельной работы обучающихся в соответствии с утвержденным учебным планом.

Индикаторы достижения компетенций считаются сформированными при достижении соответствующих им результатов обучения.

Содержание и структура дисциплины (модуля)

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
Разделы дисциплины, изучаемые в 5 семестре (очная форма).

№ разд ела	Наименование разделов	Количество часов					
		Всего	Аудиторная работа				Внеаудиторная работа СР
			Л	ЛР	ИКР	КСР	
1	2	3	4	5	6	7	8
1.	Введение в предметную область естественных языков. Основные характеристики, назначение и свойства естественных языков.	7	1	2			4
2.	Структура письменной и устной речи. Основные речевые категории. Понятие синтаксиса	7	1	2			4
3.	Морфология слов. Морфологический разбор слова. Морфологический разбор предложения.	8	2	2			4
4.	Основные синтаксические элементы. Структура предложений. Правила построения предложений.	8	2	2			4
5.	Представление предложений в виде дерева. Словно-зависимые отношения.	8	2	2			4
6.	Основные свойства проективных предложений. Деревья проективных предложений.	8	2	2			4
7.	Отрезочное представление предложений. Система составляющих. Ее свойства и построение.	8	2	2			4

8.	Инструменты для работы стекстами языка Python. Библиотеки по обработке текстовых данных. Частотный анализатор текстовой информации. Критерии анализа. Уникальность текста. Подсчет уникальности текста на основе морфологического разбора слов.	9,1	2	2	0,1	1	4
9.	Парсеры. Морфологический парсер, синтаксический парсер. Яндекс парсер. Обработка неструктурированных текстов шаблонным методом.	8,9	2	2	0,1	1	3,8
	Итого	72	16	18	0,2	2	35,8
	Зачет						
	Итого по дисциплине:	72					

Примечание: Л – лекции, КСР – контрольные и самостоятельные работы, ЛР – лабораторные занятия, СРС – самостоятельная работа студента, Д-доклад, РГЗ – расчетно-графическое задание.

Курсовые проекты или работы: не предусмотрены.

Вид аттестации: зачёт.

Основная литература

1. Лейн Хобсон, Хапки Ханес, Ховард Коул. Обработка естественного языка в действии. – СПб: Питер, 2020, – 384 с.
2. Захаров В. П., Богданова С. Ю. Корпусная лингвистика: учебник. 3-е изд., перераб. - СПб.: Изд-во С.-Петербург. ун-та, 2020. — 234 с.
3. Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тонни Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – СПб.: Питер, 2019. –368 с.
4. Алексеев В.Е. Структуры данных. Модели вычислений / В.Е. Алексеев, В.А. Таланов. - 2-е изд., исправ. - Москва: Национальный Открытый Университет «ИНТУИТ», 2016. - 248 с. : схем., ил. - (Основы информационных технологий). - Библиогр. в кн. - ISBN 5-9556-0066-3; то же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=428782>
5. Yury M. Vishnyakov, Renat Yu. Vishnyakov The Linguistic Proximity in Information Retrieval and Document Classification. // 14th IEEE International Symposium on Computational Intelligence and Informatics to be held on November 19-21, 2013 in Budapest, Hungary. p. 131-134.
6. Вишняков Ю.М., Вишняков Р.Ю. Основы вычислительной семантической интерпретации текстов естественного языка // Вестник ИМСИТа. – 2015. - №2-2. С. 13-19.
7. Вишняков Ю.М., Вишняков Р.Ю. Вычислительная семантическая интерпретация текстов научно-технического стиля // Современные наукоемкие технологии, 2016, № 12-2. – С.236-242.
8. Вишняков Ю.М., Вишняков Р.Ю. Вычислительная теория семантической интерпретации текстов научно-технического стиля естественного языка // Сборник трудов Международной научно-технической конференции «Актуальные проблемы прикладной математики, информатики и механики, Воронеж, 18–20 декабря 2017 г. – Воронеж : Издательство “Научно-исследовательские публикации”, 2017. С. 219–226. (<https://elibrary.ru/item.asp?id=32578074>).
9. Вишняков Ю.М., Вишняков Р.Ю. Об "измерении" смысловой близости текстов в рамках вычислительной теории семантической интерпретации // Сборник трудов Международной научно-технической конференции «Актуальные проблемы прикладной математики, информатики и механики, Воронеж, 17–19 декабря 2018 г. – Воронеж: Издательство “Научно-исследовательские публикации”, 2019. С. 1443–1447.

10. Yuri M. Vishnyakov, Renat Y. Vishnyakov Computational theory of semantics representation in scientific and technical texts // AMCSM_2018 IOP Publishing IOP, Conf. Series: Journal of Physics: Conf. Series 1202 (2019) 012008 doi:10.1088/1742-6596/1202/1/012008.
11. Y M Vishnyakov and R Y Vishnyakov Measurement of semantic proximity within computational theory of semantic interpretation // AMCSM_2018 IOP Publishing Journal of Physics: Conf. Series 1203 (2019) 012050 doi:10.1088/1742-6596/1203/1/012050.
12. Вишняков Ю.М., Вишняков Р.Ю. О формализации предложений научно-технического стиля естественного языка // Сборник трудов Международной научно-технической конференции «Актуальные проблемы прикладной математики, информатики и механики, Воронеж, 11–13 ноября 2019 г. – Воронеж: Издательство “Научно-исследовательские публикации”, 2019. С. 174-178.
13. Y M Vishnyakov and R Y Vishnyakov Formalization of scientific and technical sentences in Russian language // AMCSM_2019 IOP Publishing Journal of Physics: Conf. Series 1479 (2020) 012049. DOI: 10.1088/1742-6596/1479/1/012049.
14. Вишняков, Ю.М. Идентификация семантических объектов / Вишняков Р.Ю. // Актуальные проблемы прикладной математики, информатики и механики: сборник трудов Международной научной конференции, Воронеж, 7-9 декабря 2020 г. — Воронеж, 2021. С. 274-282.
15. Vishnyakov, Y M; Vishnyakov, R Y. Identification of semantic objects in information stream Journal of Physics: Conference Series; Bristol Том 1902, Изд. 1, (May 2021). DOI:10.1088/1742-6596/1902/1/012104.
16. Вишняков, Ю. М. О применении вычислительной теории семантической интерпретации к выявлению киберпреступлений /Р. Ю. Вишняков // Актуальные проблемы прикладной математики, информатики и механики: сборник трудов Международной научной конференции, Воронеж, 13-15 декабря 2021 г. — Воронеж, 2022. — С. 1526–1530.
17. Вишняков Ю. М., Вишняков Р. Ю. Машинное понимание и естественный язык / Ю. М. Вишняков, Р. Ю. Вишняков // Актуальные проблемы прикладной математики, информатики и механики: сборник трудов Международной научной конференции, Воронеж, 4-6 декабря 2023 г. — Воронеж, 2024. — С. 1375–1382.
18. Yury Vishnyakov; Renat Vishnyakov Application of Computational Theory of Semantic Interpretation to Detect Cybercrime // 2023 Applied Mathematics, Computational Science and Mechanics: Current Problems (AMCSM). Publisher: IEEE Date Added to IEEE Xplore: 21 May 2024 DOI: 10.1109/AMCSM59829.2023.10525814.
19. Вишняков Ю.М., Вишняков Р.Ю. Формализация распознавания и идентификации семантических объектов в естественно-языковых текстовых потоках// Известия ЮФУ. Технические науки, 2024, №4 – С.110-122.
20. Вишняков Ю. М., Вишняков Р. Ю. Синтез распознавателей семантических объектов // International Journal of Open Information Technologies. February 2025. Vol. 13, № 2. С. 75-80. URL: <http://injoit.org>
21. Вишняков Ю.М., Вишняков Р.Ю. «Поиск и идентификация текстов определенной семантической направленности в естественно-языковых потоках» в журнале ВАК (К1) “Современные наукоёмкие технологии” №5, 2025, DOI: 10.17513/snt.40387

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно- библиотечных системах «Лань» и «Юрайт».

Автор: профессор кафедры вычислительных технологий, д.т.н., профессор
Ю.М. Вишняков.