

АННОТАЦИЯ

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

Направление подготовки/специальности

02.03.02 Фундаментальная информатика и информационные технологии.

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

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:

- 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. 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 программистских дисциплин цикла учебного плана бакалавра.

Требования к уровню освоения дисциплины

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций: УК-4 и ПК-2.

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
1.	УК-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)	Знать грамматику и стили устной и письменной речи естественного языка	Уметь применять знания грамматики и стили устной и письменной речи естественного языка	владеть навыками осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)

№ п.п.	Индекс компетенции	Содержание компетенции (или её части)	В результате изучения учебной дисциплины обучающиеся должны		
			знать	уметь	владеть
	ПК-2	Способен проводить под научным руководством локальные исследования на основе существующих методов в конкретной области профессиональной деятельности	Подходы к исследования проблем профессиональной деятельности на основе существующих методов компьютерной лингвистики.	Проводить под научным руководством локальные исследования проблем профессиональной области на основе существующих методов компьютерной лингвистики.	Навыками проведения под научным руководством локальных исследований проблем профессиональной области на основе существующих методов компьютерной лингвистики.

Содержание и структура дисциплины (модуля)

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы дисциплины, изучаемые в _6_ семестре бакалавриата (очная форма)

№ раздела	Наименование разделов	Количество часов					
		Всего	Аудиторная работа				Внеаудиторная работа
			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. Слово-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

Курсовые проекты или работы: не предусмотрены

Вид аттестации: зачет

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.-Петербур. 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.

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