

**Государственное образовательное учреждение
«Приднестровский государственный университет им. Т.Г. Шевченко»**

Физико-технический институт

Факультет информатики и вычислительной техники

Кафедра программного обеспечения вычислительной техники

УТВЕРЖДАЮ

Директор института, доцент

Д.Н. Калошин

«30»

08

2024 г.



РАБОЧАЯ ПРОГРАММА

по дисциплине

ФТД.В.02 КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

на 2024/2025 учебный год

Направление

09.04.01 Информатика и вычислительная техника

Профиль

Информационное и программное обеспечение вычислительных систем

Квалификация

магистр

Форма обучения

очная

2023 ГОД НАБОРА

Тирасполь 2024 г.

Рабочая программа дисциплины **Компьютерная лингвистика** разработана в соответствии с требованиями Государственного образовательного стандарта ВО по направлению подготовки **09.04.01 Информатика и вычислительная техника** и основной профессиональной образовательной программы (учебного плана) по профилю подготовки **Информационное и программное обеспечение вычислительных систем**.

Составитель рабочей программы

Доцент, к.п.н.



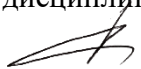
С.В. Помян

Рабочая программа утверждена на заседании кафедры *программного обеспечения вычислительной техники*

«29» августа 2024 г. протокол № 1

Зав. кафедрой, отвечающий за реализацию дисциплины,

к.т.н., доцент



С.Г. Федорченко

«29» августа 2024 г.

Зав. выпускающей кафедрой,

к.т.н., доцент



Ю.А. Столяренко

«29» августа 2024 г.

1. Цели и задачи освоения дисциплины

Целями освоения дисциплины «Компьютерная лингвистика» являются освоение методов и алгоритмов вычислительной лингвистики применительно к задачам обработки текстовой информации.

Задачами освоения дисциплины «Компьютерная лингвистика» являются формирование комплекса знаний, умений и навыков в области компьютерной (вычислительной) лингвистики, а также навыками применения существующих программных средств, реализующих эти методы и алгоритмы.

2. Место дисциплины в структуре ОПОП

Шифр дисциплины в учебном плане ФТД.В.02

Дисциплина относится к факультативной части вариативных дисциплин, формируемой участниками образовательных отношений учебного плана направления 09.04.01 Информатика и вычислительная техника в соответствии с Государственным образовательным стандартом ВО.

Общая трудоемкость дисциплины составляет 2 зачетные единицы, 72 часа.

3. Требования к результатам освоения дисциплины:

Изучение дисциплины направлено на формирование компетенций, приведенных в таблице ниже

Категория (группа) компетенций	Код и наименование	Код и наименование индикатора достижения компетенции
<i>Обязательные профессиональные компетенции выпускников и индикаторы их достижения</i>		
	ПК-1. Способен управлять развитием БД	ИД-1ПК-1 Знать: способы управления развитием БД ИД-2 ПК-1 Уметь: управлять развитием БД ИД-3ПК-1 Владеть: навыками управления развитием БД

4. Структура и содержание дисциплины

4.1. Распределение трудоемкости в з.е./часах по видам аудиторной и самостоятельной работы студентов по семестрам:

Форма обучения	Семестр (оч.ф), Курс (з.ф)	Трудоемкость, з.е./часы	Количество часов					Форма контроля
			В том числе				Самостоятельная работа (СР)	
			Аудиторных					
			Всего	Лекций (Л)	Практических (ПЗ)	Лабораторных занятий (ЛЗ)		
Очная	3	2/72	32	16	-	16	40	Зачет
	Итого:	2/72	32	16	-	16	40	Зачет

4.2. Распределение видов учебной работы и их трудоемкости по разделам дисциплины

№ Раз- дела	Наименование раздела	Количество часов				
		Всего	Аудиторная работа			СР
			Л	ПЗ	ЛЗ	
		оч.ф	оч.ф	оч.ф	оч.ф	оч.ф
1	Компьютерная морфология и син- таксис	12	2	-	2	8
2	Компьютерное представление зна- чений	14	2	-	2	10
3	Корпусная лингвистика	18	4	-	4	10
4	Информационный поиск. Извлече- ние информации	16	4	-	4	8
5	Анализ тональности. Авторство тек- ста	12	4	-	4	4
Подготовка и сдача зачета		-	-	-	-	-
Итого		72	16	-	16	40

4.3. Тематический план по видам учебной деятельности

Лекции

№ п/п	Номер раздела дисциплины	Объем часов	Тема лекций	Учебно- наглядные пособия
		±		
Компьютерная морфология и синтаксис				
1	1	2	Предмет вычислительной лингвистики	Презентация
Итого по разделу часов:		2		
Компьютерное представление значений				
2	2	2	Компьютерное представление значений	Презентация
Итого по разделу часов:		2		
Корпусная лингвистика				
3	3	2	Инструментарий вычислительной лингвистики. Словари.	Презентация
4	3	2	Корпуса текстов. Корпуса русского языка	Презентация
Итого по разделу часов:		4		
Информационный поиск. Извлечение информации				
5	4	2	Автоматический анализ текста (ААТ).	Презентация
6	4	2	Классификация документов.	Презентация
Итого по разделу часов:		4		
Анализ тональности. Авторство текста				
7	5	2	Анализ тональности.	Презентация
8	5	2	Авторство текста	Презентация
Итого по разделу часов:		4		
ИТОГО:		16		

Практические (семинарские) занятия

Учебным планом не предусмотрены.

Лабораторные занятия

№ п/п	Номер раздела дисциплины	Объем часов	Тема лабораторных занятий	Учебно- наглядные пособия
		оч. ф		
Компьютерная морфология и синтаксис				
1	1	2	Компьютерная морфология. Синтаксические анализаторы	Эл. вар. ЛР
Итого по разделу часов:		2		
Компьютерное представление значений				
2	2	2	Модели представления знаний в компьютер- ной семантике	Эл. вар. ЛР
Итого по разделу часов:		2		

Корпусная лингвистика				
3	3	2	Основные свойства корпуса. Виды корпусов	Эл. вар. ЛР
4	3	2	Корпусы русского языка. НКРЯ, OpenCorpora	Эл. вар. ЛР
Итого по разделу часов:		4		
Информационный поиск. Извлечение информации				
5	4	2	Информационный поиск.	Эл. вар. ЛР
6	4	2	Извлечение информации	Эл. вар. ЛР
Итого по разделу часов:		4		
Анализ тональности. Авторство текста				
7	5	2	Анализ тональности.	Эл. вар. ЛР
8	5	2	Авторство текста	Эл. вар. ЛР
Итого по разделу часов:		2		
ИТОГО:		16		

Самостоятельная работа обучающегося по очной форме обучения

Раздел дисциплины	№ п/п	Тема и вид самостоятельной работы обучающегося	Трудоемкость (в часах)
Компьютерная морфология и синтаксис			
Раздел 1	1.	Тема: Компьютерная морфология и синтаксис СРС №1:- работа обучающихся с лекционным материалом, - подготовка презентации по результатам поиска и анализа литературных и электронных источников.	4
	2.	Тема: Подходы к анализу синтаксических структур. Синтаксические анализаторы СРС №2:- работа обучающихся с теоретическим материалом, - подготовка обзора и сравнительной характеристики подходов к анализу.	4
Итого по разделу часов			8
Компьютерное представление значений			
Раздел 2	1	Тема: Компьютерное представление значений СРС №3:- работа обучающихся с лекционным материалом, - подготовка презентации по результатам поиска и анализа литературных и электронных источников	2
	2	Тема: Модели представления знаний в компьютерной семантике. СРС №4:- работа обучающихся с лекционным материалом, - подготовка презентации по результатам поиска и анализа литературных и электронных источников, - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных методов, алгоритмов, технологий.	2
	3	Тема: Компьютерные тезаурусы. СРС №5:- работа обучающихся с лекционным материалом, - подготовка презентации по результатам поиска и анализа литературных и электронных источников.	2

Раздел дисциплины	№ п/п	Тема и вид самостоятельной работы обучающегося	Трудоемкость (в часах)
	4	Тема: Компьютерные тезаурусы русского языка. СРС №6: - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных, методов, алгоритмов, технологий.	4
Итого по разделу часов			10
Корпусная лингвистика			
Раздел 3	1	Тема: Инструментарий вычислительной лингвистики. Словари. СРС №7:- работа обучающихся с лекционным материалом, - подготовка презентации по результатам поиска и анализа литературных и электронных источников	2
	2	Тема: Корпуса текстов. Корпуса русского языка. СРС №8:- работа обучающихся с лекционным материалом, - подготовка презентации по результатам поиска и анализа литературных и электронных источников, - поиск и анализ литературы и электронных источников информации.	2
	3	Тема: Основные свойства корпуса. СРС №9: - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных методов, алгоритмов, технологий.	2
	4	Тема: Разметка корпусов. СРС №10: - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных методов, алгоритмов, технологий.	2
	11	Тема: Корпусы русского языка. НКРЯ, OpenCorpora. СРС №11: - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных методов, алгоритмов, технологий.	2
Итого по разделу часов			10
Информационный поиск. Извлечение информации			
Раздел 4	1	Тема: Автоматический анализ текста (ААТ). Морфологический уровень СРС №12:- работа обучающихся с лекционным материалом, - подготовка презентации по результатам поиска и анализа литературных и электронных источников, - поиск и анализ литературы и электронных источников информации	2
	2	Тема: ААТ. Синтаксический уровень. Анафора и кореферентность СРС №13:- работа обучающихся с лекционным материалом, подготовка презентации по результатам поиска и анализа литературных и электронных источников, - поиск и анализ литературы и электронных источников информации, - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных методов, алгоритмов, технологий.	2

Раздел дисциплины	№ п/п	Тема и вид самостоятельной работы обучающегося	Трудоемкость (в часах)
	3	Тема: Закон Ципфа. Модель TF-IDF СРС №14:- работа обучающихся с лекционным материалом, подготовка презентации по результатам поиска и анализа литературных и электронных источников, - поиск и анализ литературы и электронных источников информации	2
	4	Тема: Классификация документов. Оценка результатов классификации. Кластеризация СРС №15:- работа обучающихся с лекционным материалом, подготовка презентации по результатам поиска и анализа литературных и электронных источников, - поиск и анализ литературы и электронных источников информации	2
Итого по разделу часов			8
Анализ тональности. Авторство текста			
Раздел 5	1	Тема: Анализ тональности текстов СРС №16:- работа обучающихся с лекционным материалом, подготовка презентации по результатам поиска и анализа литературных и электронных источников, - поиск и анализ литературы и электронных источников информации, - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных, методов, алгоритмов, технологий.	2
	2	Тема: Авторство текста. СРС №17:- работа обучающихся с лекционным материалом, подготовка презентации по результатам поиска и анализа литературных и электронных источников, - поиск и анализ литературы и электронных источников информации, - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных, методов, алгоритмов, технологий.	2
Итого по разделу часов			4
ИТОГО:			40

5. Примерная тематика курсовых проектов (работ)

Учебным планом не предусмотрены

№ п/п	Наименование учебника, учебного пособия	Автор	Год издания	Кол-во экземпляров	Электронная версия	Место размещения электронной версии
Основная литература						
1	Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. – М.: Изд-во НИУ ВШЭ, 2017. - 269 с.	Большакова Е.И. и др.	2017	-	эл. версия	Кафедра
2	Введение в компьютерную лингвистику. Учебное пособие. – СПб: НИУ ИТМО, 2013. – 72 с.	Боярский К. К.	2013	-	эл. версия	Кафедра
3	Информационные технологии в лингвистике : учеб. пособие / Л.Ю. Щипицина. – М.: ФЛИНТА : Наука, 2013. - 128 с.	Щипицина Л.Ю.	2013	-	эл. версия	Кафедра
Дополнительная литература						
4	Методы анализа компьютерных социальных сетей // Вестник НГУ. Серия: Информационные технологии. Новосибирск, 2012. Том 10, Вып. 4. С. 13–28.	Батура Т.В.	2012	-	эл. версия	Кафедра
5	Методы определения авторского стиля текстов и их программная реализация // Программные системы и вычислительные методы. М.: НБ-Медиа, 2014. № 2. С. 197–216. DOI: 10.7256/2305-6061.2014.2.11705.	Батура Т.В.	2014	-	-	http://www.nbpublish.com/library_read_article.php?id=-30093
6	В поисках синтаксической структуры: автоматический анализ русского предложения с опорой на сегментацию. М.: РГГУ. 2015. 371 с.	Кобзарева Т. Ю.	2015	-	эл. версия	Кафедра
Итого по дисциплине: 0% печатных изданий; 100 % электронных						

6.2. Программное обеспечение и Интернет- ресурсы

Программное обеспечение: *ОС Windows*, Интегрированный пакет *MS Visual Studio; SQL Server, Rational Rose 2000, UML, BP WIN*

Интернет-ресурсы

- 1) *Software Engineering Conference (Russia)* <http://www.secr.ru/>
- 2) *Software Engineering – Guide to the Software Engineering Body of Knowledge (SWE-BOK) TECHNICAL REPORT ISO/IEC TR 19759 IEEE First edition 2005-09-15.* <http://www.secr.ru/>
- 3) *CMMI® for Development, Version 1.2, CMU/SEI-2006-TR-008 ESC-TR-2006-008*

6.3. Методические указания и материалы по видам занятий

Презентации к лекционному курсу.

Варианты кейс-задач по темам по дисциплине «Компьютерная лингвистика» в электронном варианте.

7. Материально- техническое обеспечение дисциплины:

Аудитория ИВЦ, учебный кабинет.

8. Методические рекомендации по организации изучения дисциплины:

Обучающийся, изучающий дисциплину, должен, с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике.

В результате изучения дисциплины обучающийся должен знать основные определения, понятия, основные аспекты компьютерной лингвистики.

Успешное освоение курса требует самостоятельной работы обучающихся. В программе курса отведено минимально необходимое время для работы обучающихся над темой. Самостоятельная работа включает в себя:

- чтение и конспектирование рекомендованной литературы;
- проработку учебного материала (по конспектам занятий, учебной и научной литературе), подготовку ответов на вопросы, предназначенные для самостоятельного изучения, доказательство отдельных утверждений, свойств, решение задач;
- подготовка к зачету.

Руководство и контроль над самостоятельной работой обучающихся осуществляется в форме индивидуальных консультаций.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов следует обращаться за консультациями к лектору.

Технологическая карта дисциплины

Курс 2

Семестр 3

Группа ИТ23ДР68ИВ

Преподаватель – лектор Помян С.В.

Преподаватели, ведущие лабораторные, практические занятия – Помян С.В.

Наименование дисциплины/курса	Уровень образования (бакалавриат, специалитет, магистратура)	Статус дисциплины в учебном плане (А, Б)	Количество зачетных единиц	
Компьютерная лингвистика	магистратура	Б	2	
СМЕЖНЫЕ ДИСЦИПЛИНЫ ПО УЧЕБНОМУ ПЛАНУ:				
Научно-исследовательская работа, практика				
БАЗОВЫЙ МОДУЛЬ (проверка знаний и умений по дисциплине)				
Тема, задание или мероприятие текущего контроля	Виды текущей аттестации	Аудиторная или внеаудиторная	Минимальное количество баллов	Максимальное количество баллов
Презентация №1	П1	Аудиторная	5	20
Кейс-задача №1	К31	Аудиторная	7	15
Кейс-задача №2	Р1	Аудиторная	8	15
РУБЕЖНЫЙ КОНТРОЛЬ	РК		25	50
Тест №1	Т1	Аудиторная	10	20
Кейс-задача №3	К33	Аудиторная	5	15
Кейс-задача №4	К34	Аудиторная	5	15
РУБЕЖНАЯ АТТЕСТАЦИЯ	РА		25	50
		Итого	50	100