

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

**Б1.В.ДВ.01.01 ВЫЧИСЛИТЕЛЬНАЯ ЛИНГВИСТИКА**

на 2024/2025 учебный год

Направление

**09.04.04 Программная инженерия**

Профиль

**Разработка программно-информационных систем**

Квалификация

**магистр**

Форма обучения

**очная**

ГОД НАБОРА 2023

Тирасполь 2024 г.

Рабочая программа дисциплины **Вычислительная лингвистика** разработана в соответствии с требованиями Государственного образовательного стандарта ВО по направлению подготовки **09.04.04 Программная инженерия** и основной профессиональной образовательной программы (учебного плана) по профилю подготовки **Разработка программно-информационных систем**.

Составитель рабочей программы

Доцент, к.п.н.



С.В. Помян

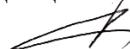
Рабочая программа утверждена на заседании кафедры *программного обеспечения вычислительной техники*

«29» августа 2024 г. протокол № 1

Зав. кафедрой, отвечающий за реализацию дисциплины,

к.т.н., доцент

«29» августа 2024 г.



С.Г. Федорченко

Зав. выпускающей кафедрой,

к.т.н., доцент

«29» августа 2024 г.



С.Г. Федорченко

### 1. Цели и задачи освоения дисциплины (модуля)

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

Задачами освоения дисциплины «Вычислительная 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 Б1.В.ДВ.01.01.

Дисциплина относится к части, формируемой участниками образовательных отношений блока Б1, дисциплина по выбору учебного плана направления 09.04.04 Программная инженерия в соответствии с Государственным образовательным стандартом ВО.

Общая трудоемкость дисциплины составляет 5 зачетных единиц, 180 часов.

### 3. Требования к результатам освоения дисциплины (модуля):

*Изучение дисциплины направлено на формирование компетенций, приведенных в таблице ниже*

Категория (группа) компетенций	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
	ПК-5. Способность выполнить постановку новых задач анализа и синтеза новых проектных решений	ИД-1 <sub>ПК-5</sub> Знает методы постановки новых задач анализа и синтеза новых проектных решений ИД-2 <sub>ПК-5</sub> Умеет использовать методы постановки новых задач анализа и синтеза новых проектных решений
	ПК-7. Способен проектировать трансляторы и интерпретаторы языков программирования.	ИД-1 <sub>ПК-7</sub> Знает методы проектирования трансляторов и интерпретаторов языков программирования ИД-2 <sub>ПК-7</sub> Умеет использовать методы проектирования трансляторов и интерпретаторов языков программирования

#### 4. Структура и содержание дисциплины (модуля)

##### 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					Самостоятельная работа (СР)	Форма контроля
			В том числе						
			Аудиторных						
			Всего	Лекций (Л)	Практических (ПЗ)	Лабораторных занятий (ЛЗ)			
Очная	3	5/180	64	32	-	32	80	Экзамен (36ч)	
	<b>Итого:</b>	5/180	64	32	-	32	80	Экзамен (36ч)	

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

№ Раздела	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 синтаксис	20	4	-	4	12
2	Компьютерное представление значений	24	4	-	8	12
3	Корпусная лингвистика	26	4	-	6	16
4	Информационный поиск. Извлечение информации	40	12	-	8	20
5	Анализ тональности. Авторство текста	34	8	-	6	20
<b>Подготовка и сдача экзамена</b>		36	-	-	-	-
<b>Итого</b>		<b>180</b>	<b>32</b>	<b>-</b>	<b>32</b>	<b>80</b>

### 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
2	1	2	Компьютерная морфология и синтаксис	Презентация
Итого по разделу часов:		<b>4</b>		
Компьютерное представление значений				
3	2	2	Компьютерное представление значений	Презентация
4	2	2	Параметры представления	Презентация
Итого по разделу часов:		<b>4</b>		
Корпусная лингвистика				
5	3	2	Инструментарий вычислительной лингвистики.	Презентация
6	3	2	Словари. Корпуса текстов.	Презентация
Итого по разделу часов:		<b>4</b>		
Информационный поиск. Извлечение информации				
7	4	2	Автоматический анализ текста (ААТ).	Презентация
8	4	2	Морфологический уровень. Синтаксический уровень	Презентация
9	4	2	ААТ. Анафора и кореферентность	Презентация
10	4	2	Классификация документов.	Презентация
11	4	2	Оценка результатов классификации	Презентация
12	4	2	Кластеризация	Презентация
Итого по разделу часов:		<b>12</b>		
Анализ тональности. Авторство текста				
13	5	2	Анализ тональности текстов.	Презентация
14		2	Алгоритмы определения тональности	Презентация
15		2	Авторство текста	Презентация
16		2	Подходы к определению авторства	Презентация
Итого по разделу часов:		<b>8</b>		
<b>ИТОГО:</b>		<b>32</b>		

#### Практические (семинарские) занятия

Учебным планом не предусмотрены.

### Лабораторные занятия

№ п/п	Номер раздела дисциплины	Объем часов	Тема лабораторных занятий	Учебно-наглядные пособия
1	1	2	Компьютерная морфология	Эл.вариант лаб.раб
2	1	2	Подходы к анализу синтаксических структур	Эл.вариант лаб.раб
Итого по разделу часов:		<b>4</b>		
3	2	2	Модели представления знаний в компьютерной семантике	Эл.вариант лаб.раб
4	2	2	Разновидности семантических представлений	Эл.вариант лаб.раб
5	2	2	Компьютерные тезаурусы	Эл.вариант лаб.раб
6	2	2	Компьютерные тезаурусы	Эл.вариант лаб.раб
Итого по разделу часов:		<b>8</b>		
7	3	2	Основные свойства корпуса.	Эл.вариант лаб.раб
8	3	2	Виды корпусов	Эл.вариант лаб.раб
9	3	2	Корпусы русского языка. НКРЯ, OpenCorpora	Эл.вариант лаб.раб
Итого по разделу часов:		<b>6</b>		
10	4	2	Информационный поиск. Поисковая лингвистика	Эл.вариант лаб.раб
11	4	2	Извлечение информации из текста на естественном языке	Эл.вариант лаб.раб
12	4	2	Распознавание сущностей.	Эл.вариант лаб.раб
13	4	2	Извлечение отношений и событий	Эл.вариант лаб.раб
Итого по разделу часов:		<b>8</b>		
14	5	2	Анализ тональности текста	Эл.вариант лаб.раб
15	5	2	Авторство текста	Эл.вариант лаб.раб
16	5	2	Формальные методы определения авторства текста	Эл.вариант лаб.раб
Итого по разделу часов:		<b>6</b>		
<b>ИТОГО:</b>		<b>32</b>		

*Самостоятельная работа обучающегося по очной форме обучения*

Раздел дисциплины	№ п/п	Тема и вид самостоятельной работы обучающегося	Трудоемкость (в часах)
<b>Компьютерная морфология и синтаксис</b>			
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.	6
	2	Тема: Подходы к анализу синтаксических структур. Синтаксические анализаторы СРС №2:- работа обучающихся с теоретическим материалом, подготовка обзора и сравнительной характеристики подходов к анализу.	6
<b>Итого по разделу часов</b>			<b>12</b>
<b>Компьютерное представление значений</b>			
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.	4
	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
<b>Итого по разделу часов</b>			<b>12</b>
<b>Корпусная лингвистика</b>			
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	2

Раздел дисциплины	№ п/п	Тема и вид самостоятельной работы обучающегося	Трудоемкость (в часах)
		и анализа литературных и электронных источников	
	3	Тема: Основные свойства корпуса. СРС №9: - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных методов, алгоритмов, технологий.	4
	4	Тема: Разметка корпусов. СРС №10: - подготовка к выполнению индивидуальных заданий кейс-задачи, подготовка ответов на вопросы по обоснованию использованных методов, алгоритмов, технологий.	4
	5	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.	4
<b>Итого по разделу часов</b>			<b>16</b>
<b>Информационный поиск. Извлечение информации</b>			
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	4
	2	Тема: ААТ. Синтаксический уровень. Анафора и ко-референтность СРС №13:- 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.	4
	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	6
	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	6
<b>Итого по разделу часов</b>			<b>20</b>
<b>Анализ тональности. Авторство текста</b>			
5	1	Тема: Анализ тональности текстов СРС №16:- работа обучающихся с лекционным мате-	8

Раздел дисциплины	№ п/п	Тема и вид самостоятельной работы обучающегося	Трудоемкость (в часах)
		риалом, 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	Тема: Авторство текста. СРС №17:- 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.	10
<b>Итого по разделу часов</b>			<b>20</b>
<b>Подготовка и сдача экзамена</b>			<b>36</b>
<b>ИТОГО:</b>			<b>116</b>

## 5. Примерная тематика курсовых проектов (работ)

Учебным планом не предусмотрены

## 6. Учебно- методическое и информационное обеспечение дисциплины (модуля)

### 6.1 Обеспеченность обучающихся учебниками, учебными пособиями

№ 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. и др. – 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. -	Щипицина Л.Ю.	2013	-	эл. версия	Кафедра

№ п/п	Наименование учебника, учебного пособия	Автор	Год издания	Кол-во экземпляров	Электронная версия	Место размещения электронной версии
	128 с.					
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	-	эл. версия	<a href="http://www.nbpublish.com/library_read_article.php?id=-30093">http://www.nbpublish.com/library_read_article.php?id=-30093</a>
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
<i>Итого по дисциплине: 0% печатных изданий; 100 % электронных</i>						

## **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 «Вычислительная 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 программной инженерии.

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 экзамену.

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 ИТ23ДР68ПИ

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	
Вычислительная лингвистика	магистратура	А	5	
<b>СМЕЖНЫЕ ДИСЦИПЛИНЫ ПО УЧЕБНОМУ ПЛАНУ:</b>				
Научно-исследовательская работа, практика				
<b>БАЗОВЫЙ МОДУЛЬ</b> (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	10
Презентация №2	П2	Аудиторная	5	10
Кейс-задача №1	К31	Аудиторная	7	15
Кейс-задача №2	Р1	Аудиторная	8	15
<b>РУБЕЖНЫЙ КОНТРОЛЬ</b>	<b>РК</b>		<b>25</b>	<b>50</b>
Тест №1	Т1	Аудиторная	10	20
Кейс-задача №3	К33	Аудиторная	5	15
Кейс-задача №4	К34	Аудиторная	5	15
<b>РУБЕЖНАЯ АТТЕСТАЦИЯ</b>	<b>РА</b>		<b>25</b>	<b>50</b>
		<b>Итого</b>	<b>50</b>	<b>100</b>