

Министерство науки и высшего образования  
Российской Федерации

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Донецкий государственный университет»

Филологический факультет

Кафедра общего, славянского и прикладного языкознания имени Е. С. Отина



УТВЕРЖДАЮ

проректор

П. А. Машаров

«29» марта 2024 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**АВТОМАТИЧЕСКАЯ ОБРАБОТКА ЕСТЕСТВЕННЫХ ЯЗЫКОВ**

Укрупненная группа направлений подготовки	45.00.00	Языкознание	и
Программа высшего образования		литературоведение	
Направление подготовки	Программа бакалавриата		
	45.03.03	Фундаментальная и прикладная лингвистика	
Профиль подготовки	Фундаментальная	и	прикладная лингвистика
Квалификация	Бакалавр		
Форма обучения	Очная		

Рабочая программа адаптирована для лиц  
с ограниченными возможностями здоровья и инвалидов

Донецк 2024



Рабочая программа дисциплины «Автоматическая обработка естественных языков» для обучающихся по направлению подготовки 45.03.03 Фундаментальная и прикладная лингвистика (Профиль: Фундаментальная и прикладная лингвистика), составлена на основании Федерального государственного образовательного стандарта высшего образования – бакалавриат по направлению подготовки 45.03.03 Фундаментальная и прикладная лингвистика, утвержденного приказом Министерства образования и науки Российской Федерации от 24 апреля 2018 г. № 323 (с изм. и доп.), Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 06 апреля 2021 г. № 245 (с изм. и доп.), в соответствии с учебным планом, утвержденным Ученым советом ФГБОУ ВО «ДонГУ» для набора 2024 года.

Разработчик:

старший преподаватель кафедры общего,  
славянского и прикладного языкознания  
имени Е. С. Отина



С. А. Могила

Рабочая программа одобрена на заседании кафедры общего, славянского и прикладного языкознания имени Е. С. Отина.

Протокол от 26.03.2024 г. № 9.

И. о. заведующего кафедрой



Н. А. Ярошенко

СОГЛАСОВАНО:

Декан филологического факультета  
28.03.2024 г.



Н. А. Ярошенко

Учебно-методическая комиссия филологического факультета.

Протокол от 27.03.2024 г. № 3.

Председатель



С. В. Руденко

Руководитель основной профессиональной  
образовательной программы,  
канд. филол. наук, доц.  
26.03.2024 г.



Н. А. Ярошенко

## 1. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

1.1. Требования к предварительной подготовке обучающихся, предшествующие и сопутствующие дисциплины, на которых основывается изучение данной дисциплины программы бакалавриата: Основы создания компьютерных словарей, Методика преподавания информатики.

1.2. Дисциплины, для которых освоение данной дисциплины необходимо как предшествующее: Компьютерные технологии в филологии, Компьютерный анализ текста, Квантитативная лингвистика.

## 2. ОПИСАНИЕ ДИСЦИПЛИНЫ

### 2.1. Общая характеристика

<i>Наименование показателя</i>	<i>Значение показателя</i>
Название образовательной программы	45.03.03 Фундаментальная и прикладная лингвистика
Шифр и название в соответствии с учебным планом	Б1.Б.31 Автоматическая обработка естественных языков
Часть образовательной программы	Базовая (обязательная) часть
Количество зачетных единиц	9

### 2.2. Распределение часов по форме и периоду обучения

Форма обучения	курс	семестр	Общее количество часов					Форма контроля
			лекционных	лабораторных	практических	самостоятельной работы + контроль	всего	
Очная	2	3,4	49	115		160	324	зачет, экзамен

### 3. ЦЕЛИ ДИСЦИПЛИНЫ

Ознакомить студентов с современным состоянием и достижениями в области автоматической обработки текста, распознавания, анализа и синтеза речи в звуковой и текстовой форме; научить пользоваться отечественными и зарубежными лингвистическими программными продуктами (экспертными, диалоговыми, учебными, информационными поисковыми системами).

### 4. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ КОМПОНЕНТА ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ, ИХ ИНДИКАТОРЫ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Компетенции	Индикаторы	Результаты обучения
ОПК-2. Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур	ОПК-2.1. Демонстрирует знание основ математических дисциплин, необходимых для формализации лингвистических знаний.	ОПК-2.1.1. Знает основы математических дисциплин, необходимые для формализации лингвистических знаний.
		ОПК-2.1.2. Умеет соотносить задачи формализации лингвистических знаний с основами математических знаний.
		ОПК-2.1.3. Владеет математическими методами в процедурах анализа и синтеза лингвистических структур.
ПК-2. Способен вести мониторинг информационных массивов и готовить на этой основе аналитические материалы.	ПК-2.1. Понимает, критически анализирует и использует базовую информацию в предметной области.	ПК-2.1.1. Знает, критически анализирует и использует базовую информацию в предметной области.
		ПК-2.1.2. Демонстрирует умение к составлению обзоров, аннотаций, рефератов и библиографии по тематике проводимых исследований

### 5. ПРОГРАММА ДИСЦИПЛИНЫ

Название темы	Краткое содержание темы (вопросы темы)
<b>Тема 1. Язык как знаковая система</b>	<b>Язык как знаковая система.</b> Язык – средство передачи информации. Формы коммуникации. Естественные и формальные языки. Теория автоматической обработки текста. Роль компьютера как инструмента познания действительности. Основные направления компьютерной лингвистики.
<b>Тема 2. Техническая обработка текста.</b>	<b>Техническая обработка текста (word processing).</b> Система автоматического распознавания текстов. Алгоритмы распознавания. Сканирование текста.
<b>Тема 3. Автоматическая обработка текста как совокупности слов.</b>	<b>Автоматическая обработка текста как совокупности слов (word matching).</b> Автоматическое индексирование. Автоматическая классификация документов. Создание словарей, конкордансов, частотных словарей.
<b>Тема 4. Система OCR.</b>	<b>Система OCR оптического распознавания символов.</b>

	Основные принципы работы системы. Порядок распознавания текста. Система оптического распознавания текстов ABBYY FineReader
<b>Тема 5. Автоматизация лингвистического анализа текста.</b>	<b>Автоматизация лингвистического анализа текста.</b> Морфологический анализ текста с использование приложений MS Office. Синтаксический анализ текста с использование приложений MS Office.
<b>Тема 6. Автоматизация письменного перевода текста.</b>	<b>Автоматизация письменного перевода текста.</b> Основные виды письменного перевода. Основные подходы к переводу. Алгоритм автоматизированного перевода письменного текста. Перевод письменного текста в среде Microsoft Office. Использование автоматизированных переводчиков для письменного перевода текста. Перевод и локализация сайтов
<b>Тема 7. Компрессия письменного текста.</b>	<b>Компрессия письменного текста.</b> Информационная компрессия. Основные варианты компрессии письменного текста. Семиотические и коммуникативные способы информационной компрессии. Автоматическое реферирование и аннотирование. Алгоритм построения реферата и аннотации. Системы автоматического реферирования и аннотирования.
<b>Тема 8. Тезаурусы.</b>	<b>Тезаурусы.</b> История создания тезаурусов. Основные понятия. Основные отношения в тезаурусе. Основные функции и задачи тезауруса. Использование тезауруса в приложениях MS Office.
	<b>Содержательный модуль 2. Синтез и распознавания речи в компьютерных системах.</b>
<b>Тема 9. Синтез речи.</b>	<b>Синтез речи.</b> Основные понятия синтеза речи. История синтеза речи. Области применения синтеза речи. Способы синтеза речи. Упрощенная технология синтеза речи.
<b>Тема 10. Распознавание речи.</b>	<b>Распознавание речи.</b> История систем распознавания речи. Область применения систем распознавания речи. Звуковое оборудование компьютера. Системы распознавания речи.
<b>Тема 11. Шумоочистка речевых сигналов.</b>	<b>Шумоочистка речевых сигналов.</b> Основные задачи. Общая схема шумоочистки. Программные комплексы редактирования, анализа и шумоочистки звуковых (речевых) сигналов.
<b>Тема 12. Оценочная лексика.</b>	<b>Оценочная лексика.</b> Основные группы стилистической окраски лингвистической единицы. Функция воздействия и степень выразительности языковых средств. Фоносемантическая оценка словоформ и письменного текста. Программа ВААЛ.

6. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ  
6.1. Форма обучения – очная, курс – 2, семестр – 3

Наименования разделов и тем	Количество часов				
	Лекц.	Лабор.	Практ.	СРС+К	Всего
<b>Тема 1. Язык как знаковая система.</b>	2	4		10	16
<b>Тема 2. Техническая обработка текста.</b>	2	6		10	18
<b>Тема 3. Автоматическая обработка текста.</b>	2	6		10	18
<b>Тема 4. Система OCR.</b>	2	4		10	16
<b>Тема 5. Автоматизация лингвист. анализа текста.</b>	2	8		10	20
<b>Тема 6. Автоматизация письм-го перевода текста.</b>	2	8		10	20
<b>Тема 7. Компрессия письмен. текста.</b>	3	8		10	21
<b>Тема 8. Тезаурусы.</b>	2	7		10	19
<b>Итого за 3 семестр:</b>	<b>17</b>	<b>51</b>		<b>80</b>	<b>148</b>

6.2. Форма обучения – очная, курс – 2, семестр – 4

Наименования разделов и тем	Количество часов				
	Лекц.	Лабор.	Практ.	СРС+К	Всего
<b>Тема 9. Синтез речи.</b>	8	28		20	51
<b>Тема 10. Распознавание речи.</b>	8	12		20	35,8
<b>Тема 11. Шумоочистка речевых сигналов</b>	8	12		20	40
<b>Тема 12. Оценочная лексика.</b>	8	12		20	40
<b>Итого за 4 семестр:</b>	<b>32</b>	<b>64</b>		<b>80</b>	<b>166</b>

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ (СРЕДСТВА) ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Контрольные вопросы

1. Основные понятия языка как знаковой системы для порождения, передачи и хранения информации.
2. Формальные языки. Основное отличие формальных языков от естественных.
3. Компьютерная лингвистика как новый этап в развитии теоретической и прикладной лингвистики.
4. Основные этапы автоматической обработки письменного текста.
5. Техническая обработка текста (word processing). Основные этапы.
6. Основные принципы работы системы OCR (Optical Character Recognition).
7. Общая характеристика системы оптического распознавания текстов ABBYY FineReader.
8. Основные этапы работы системы оптического распознавания текстов ABBYYFineReader.
9. Поиск текста в документе. Работа с панелью Навигация.
10. Замена текста в документе. Использование окна «Найти и Заменить».
11. Основные виды письменного перевода.
12. Критерий «прозрачности перевода». Основные понятия.
13. Основные разновидности письменного перевода.
14. Виды переводов.
15. Автоматизация лингвистического анализа текста.
16. Автоматизация морфологического анализа.
17. Автоматизация синтаксического анализа.
18. Тезаурусы. Основные понятия.
19. Основные функции тезаурусов.
20. Компрессия текста. Основные понятия.
21. Компрессия текста.
22. Основные типы компрессированных вариантов представления текста.
23. Компрессия текста с использованием набора ключевых слов.

24. Семиотические способы информационной компрессии.
25. Коммуникативные способы информационной компрессии.
26. Этапы построения реферата человеком.
27. Компрессия текста. Основные понятия.
28. Методы автоматического реферирования и аннотирования.
29. Порядок составления реферата или аннотации при помощи компьютера.
30. Характеристика различных групп методов автоматического реферирования и аннотирования текстов.
31. Статистическая группа методов автоматического реферирования и аннотирования текстов.
32. Модель линейных весовых коэффициентов при автоматическом реферировании и аннотировании.
33. Модель реферирующих систем с использованием базы знаний при автоматическом реферировании и аннотировании.
34. Алгоритм автоматического построения реферата и аннотации письменного текста.
35. Использование программы ОРФО для проведения лингвистического анализа текста.

## 8. РАСПРЕДЕЛЕНИЕ БАЛЛОВ, КОТОРЫЕ ПОЛУЧАЮТ ОБУЧАЮЩИЕСЯ

Общая оценка знаний обучающихся по дисциплине проводится по 100-балльной шкале исходя из максимума, приведенного в таблице ниже. Организационно-учебная работа в аудитории оценивается на основе таких критериев как посещаемость занятий, своевременное и качественное выполнение домашних заданий, активность во время проведения лекционных и лабораторных занятий.

### 8.1. Семестр 3

Номера разделов	Виды работ	Максимальное количество баллов
1-8	Организационно-учебная работа в аудитории	40
	Самостоятельная работа	10
	Контрольные работы по практике	
	Контрольная работа по теоретическому материалу	
ИТОГО		
Зачет		50
Общий итог за семестр		100

### 8.1. Семестр 4

Номера разделов	Виды работ	Максимальное количество баллов
1-8	Организационно-учебная работа в аудитории	40
	Самостоятельная работа	10
	Контрольные работы по практике	
	Контрольная работа по теоретическому материалу	
ИТОГО		
Экзамен		50
Общий итог за семестр		100



## Соответствие баллов оценке

Количество баллов из 100	ECTS	Оценка по пятибалльной шкале	
		Экзамен, дифференцированный зачет	Зачет
90-100	A	отлично	зачтено
80-89	B	хорошо	зачтено
75-79	C		зачтено
70-74	D	удовлетворительно	зачтено
60-69	E		зачтено
35-59	FX	неудовлетворительно	не зачтено
0-34	F		не зачтено

### 9. ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ДЛЯ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ И ИНВАЛИДОВ

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- 1) для слепых и слабовидящих:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
  - письменные задания оформляются увеличенным шрифтом.
- 2) для глухих и слабослышащих:
  - лекции оформляются в виде электронного документа;
  - письменные задания выполняются на компьютере в письменной форме;
  - экзамен проводится в письменной форме на компьютере; возможно проведение в форме тестирования.
- 3) для лиц с нарушениями опорно-двигательного аппарата:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере;
  - экзамен и зачет проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- 1) для слепых и слабовидящих:
  - в печатной форме увеличенным шрифтом;
  - в форме электронного документа;
- 2) для глухих и слабослышащих:
  - в печатной форме;
  - в форме электронного документа.
- 3) для обучающихся с нарушениями опорно-двигательного аппарата:
  - в печатной форме;

- в форме электронного документа.

## 10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОГО ПРОЦЕССА

Для проведения лабораторных занятий требуется аудитория, оборудованная меловой или маркерной доской, мультимедийный проектор и экран, ноутбук, комплект учебной мебели для студентов, рабочее место преподавателя. Выход в Интернет проводной или с использованием Wi-Fi.

Для самостоятельной работы используются текстовые и электронные ресурсы Научной библиотеки университета и других электронных библиотечных баз данных.

При изучении дисциплины применяются электронное обучение и дистанционные образовательные технологии.

С использованием ресурсов платформы дистанционного образования осуществляется текущий контроль знаний обучающихся на основе тестирования и проверки результатов самостоятельной работы.

## 11. РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

### 11.1. Основная литература

1. Баранов А. Н. Введение в прикладную лингвистику: Учеб. пособие / А. Н. Баранов; Моск. гос. ун-т им. М. В. Ломоносова, Филол. фак. – Москва, Эдиториал УРСС, 2003. – 358 с.
2. Кузнецов, С. Д. Базы данных: модели и языки: учеб. пособ. для студентов вузов, обучающихся по специальности " Прикладная математика и информатика" и "Информационные технологии" / С. Д. Кузнецов. - М.: Бином, 2008. - 720 с.
3. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы: [учеб. пособие для лингв. фак. вузов] / Н. Н. Леонтьева. - М.: Academia, 2006. - 303 с.
4. Введение в прикладную лингвистику: учебное пособие / Е. П. Соснина. – 2-е изд., испр. и доп. – Ульяновск : УлГТУ, 2012. – 110 с.

### 11.2. Дополнительная литература

1. Генерация тонального контура фразы в системах автоматического синтеза речи // Труды международного семинара Диалог'2000 по компьютерной лингвистике и приложениям. Протвино, 2000. [С. 211-219]. [www.dialog-21.ru/digest/2000/articles/krivnova/](http://www.dialog-21.ru/digest/2000/articles/krivnova/)
2. Маслова, В. А. Современные направления в лингвистике: учеб. пособие для студентов вузов, обучающихся по направлению "Филол. образование" / В. А. Маслова. - Москва : Академия, 2008. - 264 с.
3. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи. Минск, 2008. [pselab.ru/Books/Lobanov\\_Cirulnik\\_2008.pdf](http://pselab.ru/Books/Lobanov_Cirulnik_2008.pdf)

## 12. ИНФОРМАЦИОННЫЕ РЕСУРСЫ

1. Грамота.Ру: справочно-информационный портал «Русский язык» <http://www.gramota.ru/>.
2. Культура письменной речи <http://www.gramma.ru/>.
3. Справочная служба русского языка <http://spravka.gramota.ru>.

## 13. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

1. Windows 7 PRO (корпоративная лицензия ДонГУ № 46484614)
2. Microsoft Office (корпоративная лицензия ДонГУ № 46472919)
3. Microsoft Visual Studio (лицензия программы Dream Spark для высших учебных заведений)
4. Антивирус Касперского, Adobe Acrobat Reader, xPDF (лицензии GPL, Apache, BSD для свободного программного обеспечения).