

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

### КОМПЬЮТЕРНЫЙ АНАЛИЗ ТЕКСТА

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 «Компьютерный анализ текста» 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:

доцент кафедры общего, славянского и  
прикладного языкознания имени Е. С. Отина,  
канд. филол. наук



Э. А. Титаренко

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: Общая и русская лексикология, Общая и русская фонетика, Общая и русская морфология, Общий и русский синтаксис.

1.2. Дисциплины, курсовые работы и практики, для которых освоение данной дисциплины необходимо как предшествующее: Производственная практика: преддипломная, подготовка ВКР; Подготовка к процедуре защиты и защита выпускной квалификационной работы.

## 2. ОПИСАНИЕ ДИСЦИПЛИНЫ

### 2.1. Общая характеристика

Наименование показателя	Значение показателя
Название образовательной программы	45.03.03 Фундаментальная и прикладная лингвистика (Профиль: Фундаментальная и прикладная лингвистика)
Шифр и название в соответствии с учебным планом	Б1.В.ДВ.3.2. Компьютерный анализ текста
Часть образовательной программы	Вариативная часть
Количество зачетных единиц / всего часов	2 / 72

### 2.2. Распределение часов по формам и периодам обучения

Форма обучения	курс	семестр	Общее количество часов					Форма контроля
			лекционных	лабораторных	практических	самостоятельной работы + контроль	всего	
Очная	4	8	18	–	18	36	72	зачет

## 3. ЦЕЛИ ДИСЦИПЛИНЫ

Цель дисциплины — получение новых знаний о видах компьютерного анализа текстов и совершенствование профессиональных компетенций в области прикладной лингвистики.

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

Компетенции	Индикаторы	Результаты обучения
ОПК-1. Способен применять в профессиональной деятельности основные понятия и категории	ОПК-3.1. Использует знания об истории, современном состоянии и перспективах развития лингвистики и ее	ОПК-3.1.1. Знает виды компьютерного анализа текста; программы, используемые для компьютерного анализа текстов. ОПК-3.1.2. Умеет использовать

современной лингвистики.	конкретной области в профессиональной, в том числе педагогической, деятельности.	навыки компьютерного анализа текстов в профессиональной деятельности.
--------------------------	--	---

## 5. ПРОГРАММА ДИСЦИПЛИНЫ

Название темы	Краткое содержание темы (вопросы темы)
Раздел 1.	
Тема 1. Виды компьютерного анализа текста. Графематический анализ.	Задачи и виды компьютерного анализа текста Выделение из текста синтаксических и структурных единиц: абзацы, предложения, слова и знаки препинания
Тема 2. Морфологический анализ	Приведение слов к нормальной форме и назначение им соответствующих параметров: часть речи, род, падеж, время и т. д.
Тема 3. Предсинтаксический анализ	Разделение одной лексической единицы на несколько синтаксических и наоборот, а также проведение синтаксической сегментации. Объединение отдельных лексических единиц, разделение одной лексической единицы, свёртка числительных и синтаксическая сегментация.
Тема 4. Синтаксический и постсинтаксический анализ текста	Определение синтаксической роли слов и синтаксических связей между ними. Уточнение смысла слов и нормализация синтаксического дерева. Сведение конструкций, выражающих одну и ту же мысль разными способами на разных языках, в один и тот же вид.
Тема 5. Семантический анализ.	Выявление смысла слов и формирование между ними семантических связей, построение семантических сетей и онтологий
Тема 6. Автоматическое аннотирование текстов	Вытягивающие алгоритмы автоматического аннотирования текстов. Генерирующие алгоритмы автоматического аннотирования текстов.
Тема 7. Автоматическое извлечение ключевых слов	Статистическая модель автоматического извлечение ключевых слов. Гибридная модель автоматического извлечение ключевых слов.

## 6. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

### 6.1. Форма обучения – очная, курс – 4, семестр – 8

Наименования разделов и тем	Количество часов				
	Лекц.	Лабор.	Практ.	СРС+К	Всего
Тема 1. Виды компьютерного анализа текста. Графематический анализ.	2		2	5	9
Тема 2. Морфологический анализ	2		2	5	9
Тема 3. Предсинтаксический анализ	2		2	6	10

Тема 4. Синтаксический и постсинтаксический анализ текста	2		2	6	10
Тема 5. Семантический анализ.	2		2	5	9
Тема 6. Автоматическое аннотирование текстов	4		4	5	13
Тема 7. Автоматическое извлечение ключевых слов	4		4	4	12
ИТОГО ПО КОМПОНЕНТУ ОПОП	18	—	18	36	72

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

### 7.1. Контрольные вопросы

#### Раздел 1

Задачи и виды компьютерного анализа текста.

Выделение из текста синтаксических и структурных единиц: абзацы, предложения, слова и знаки препинания.

Предсинтаксический анализ.

Определение синтаксической роли слов и синтаксических связей между ними.

Постсинтаксический анализ текста

Семантический анализ.

Автоматическое аннотирование текстов.

Статистическая модель автоматического извлечение ключевых слов.

Гибридная модель автоматического извлечение ключевых слов.

### 7.2. Темы письменных работ (типы заданий)

Выполнить частотный анализ слов и словосочетаний.

Выполнить поиск по ключевым словам.

Выполнить поиск по контексту.

Контрольная работа по проверке теоретических знаний – по всем темам, с использованием указанных выше контрольных вопросов.

В случае ведения учебного процесса с использованием электронного обучения и дистанционных образовательных технологий, содержание билета может отличаться от приведенного.

## 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 практических занятий (участие в обсуждении текущего и пройденного материала, решение задач и т.п.).

### 8.1. Семестр 8

Номера разделов	Виды работ	Максимальное количество баллов
1	Организационно-учебная работа в аудитории	20

	Самостоятельная работа	10
	Контрольная работа	30
Зачет		40
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 зачёт 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

Учебные занятия проводятся в 1-м учебном корпусе ДонГУ (г. Донецк, ул. Университетская, 24). Для проведения лекционных и практических 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, учебно-методическое обеспечение, представленное в учебно-методическом кабинете Главного корпуса (ауд.405).

Обучающиеся имеют возможность использовать учебные материалы по дисциплине, размещенные на платформе Moodle Центра дистанционного образования ФГБОУ ВО «ДонГУ». 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. Дюжева А. Н. Информационные технологии в лингвистике. – 2021.
  2. Большакова Е. И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. – 2011.
  3. Аслонов Ш. Ш. и др. Компьютерная лингвистика и филология: проблемы и решения //Гуманитарный трактат. – 2020. – №. 84. – С. 17-19.
  4. Бутусова, А. С. Машинный и автоматизированный перевод : учебное пособие / А.С. Бутусова, Ю. В. Бец ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог: Издательство Южного федерального университета, 2021. – 106 с
- Дополнительная
- Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004
- Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011.
- Хименко, В.И. Случайные данные: структура и анализ / В.И. Хименко. – Москва : Техносфера, 2017. – 424 с.

Шереметьева С. О., Осминин П. Г. Методы и модели автоматического извлечения ключевых слов //Вестник ЮжноУральского государственного университета. Серия: Лингвистика. – 2015. – Т. 12. – №. 1. – С. 76-81

## 12. ИНФОРМАЦИОННЫЕ РЕСУРСЫ

1. **Национальная электронная библиотека (НЭБ):** федеральная государственная информационная система / Министерство Культуры РФ; Российская государственная библиотека. – Москва, 2019- . – URL: <https://rusneb.ru/> (дата обращения: 01.09.2023). – Режим доступа: свободный, подписка. Необходима установка программного обеспечения. – Текст: электронный.
2. **eLIBRARY.RU:** научная электронная библиотека: сайт. – Москва, 2000- . – URL: <https://elibrary.ru> (дата обращения: 01.09.2023). – Режим доступа: для авторизов. пользователей. –Текст: электронный.
3. Научная электронная библиотека **«КиберЛенинка»:** сайт / Ассоциация «Открытая наука». – Москва, 2014- . – URL: <https://cyberleninka.ru/>. – Режим доступа: свободный. – Текст: электронный.
4. Электронно-библиотечная система **«Лань»:** [сайт]. – URL: <https://e.lanbook.com> (дата обращения: 01.09.2023). – Режим доступа: для авторизов. пользователей. – Текст: электронный.
5. **ЭБС Юрайт:** электронная библиотечная система: сайт. – Москва, 2013. – URL: <https://biblio-online.ru> (дата обращения: 01.09.2023). – Режим доступа: для авторизов. пользователей. – Текст: электронный.
6. **Электронно-библиотечная система ДонГУ:** сайт / ФГБОУ ВО «ДонГУ». – Донецк, 2016- . – URL: <http://library.donnu.ru/> (дата обращения: 01.09.2023). – Режим доступа: свободный. – Текст: электронный.
7. **Электронный каталог** Научной библиотеки ДонГУ: раздел сайта / НБ ДонГУ. – Текст: электронный // ЭБС ДонГУ: сайт. – URL: <http://library.donnu.ru/catalog/> (дата обращения: 01.09.2023). – Режим доступа: поиск свободный, электронные документы – для пользователей ДонГУ.
8. **Электронный архив ДонГУ:** раздел сайта / НБ ДонГУ. – Текст: электронный // ЭБС ДонГУ: сайт. – URL: <http://repo.donnu.ru/> (дата обращения: 01.09.2023). – Режим доступа: свободный.

## 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).



