

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
«ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»



УТВЕРЖДАЮ:

Проректор по научно-педагогической работе

А.В. Левшов

(подпись)

« 02 » 01 2017 года

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

Компьютерная лингвистика

(наименование дисциплины согласно учебному плану)

Направление  
подготовки:

09.03.01 «Информатика и вычислительная техника»  
(код и наименование направления / специальности)

Профиль:

«Программное обеспечение средств вычислительной  
техники»

(наименование профиля / магистерской программы / специализации)

Программа:

бакалавриат

(бакалавриат, магистратура, специалитет)

Форма обучения:

очная

(очная, заочная, очно-заочная)

Форма обучения	очная
Семестр(ы)	3
Общая трудоёмкость в з.е./часах	6,0/216
Аудиторные занятия (час.), в том числе	85
Лекции (час.)	34
Практические (семинарские) занятия (час.)	17
Лабораторные работы (час.)	34
Самостоятельная работа (час.), в том числе	77
Курсовой проект(работа) (семестр/час.)	3(36 час)
Индивидуальное задание (кол./час.)	-
Форма промежуточной аттестации (экзамен(зачёт), час.)	экзамен (54 часа)

Донецк, 2017 г.

Рабочая программа дисциплины «Компьютерная лингвистика» составлена в соответствии с учебным планом по направлению подготовки 09.03.01 «Информатика и вычислительная техника», профиль подготовки «Программное обеспечение средств вычислительной техники» для 2017 года приёма.

Составитель: Мальчева Р.В., к.т.н., доц., профессор кафедры компьютерной инженерии.

Рабочая программа **рассмотрена и утверждена** на заседании выпускающей кафедры компьютерной инженерии.

Протокол от «14» декабря 2016 года № 3

Заведующий кафедрой \_\_\_\_\_ Аноприенко А.Я.  
(подпись)

Рабочая программа **одобрена учебно-методической комиссией** ДонНТУ по направлению подготовки «Информатика и вычислительная техника»

Протокол от «14» декабря 2016 года № 2

Председатель \_\_\_\_\_ Аноприенко А.Я.  
(подпись)

Рабочая программа **продлена** для 2017 года приёма на заседании кафедры компьютерной инженерии.

Протокол от «20» июня 2017 года № 10

Заведующий кафедрой \_\_\_\_\_ Аноприенко А.Я.  
(подпись)

Согласовано с выпускающей кафедрой компьютерной инженерии

Заведующий кафедрой \_\_\_\_\_ Аноприенко А.Я.  
(подпись) (Ф.И.О.)

Рабочая программа **продлена** для 2018 года приёма на заседании кафедры компьютерной инженерии.

Протокол от «31» 08 2018 года № 1

Заведующий кафедрой \_\_\_\_\_ Аноприенко А.Я.  
(подпись)

Согласовано с выпускающей кафедрой компьютерной инженерии

Заведующий кафедрой \_\_\_\_\_ Аноприенко А.Я.  
(подпись) (Ф.И.О.)

Рабочая программа **продлена** для 2019 года приёма на заседании кафедры компьютерной инженерии.

Протокол от «30» 08 2019 года № 1

Заведующий кафедрой \_\_\_\_\_ Аноприенко А.Я.  
(подпись)

Согласовано с выпускающей кафедрой компьютерной инженерии

Заведующий кафедрой \_\_\_\_\_ Аноприенко А.Я.  
(подпись) (Ф.И.О.)

## **1. ОБЪЕКТ, ЦЕЛЬ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

Цель дисциплины - знакомство с основными проблемами, современными методами и подходами, инструментами и технологиями в области компьютерной лингвистики.

Задачи дисциплины:

- знакомство с основными проблемами, базовыми алгоритмами, математическими методами, основными инструментами и технологиями в области автоматической обработки естественного языка;
- приобретение умения представлять в алгоритмическом виде процессы анализа и синтеза текста;
- развитие способности осуществлять лингвистическую обработку текстов в производственно-практических целях, разрабатывать и совершенствовать электронные языковые ресурсы, лингвистические компоненты информационных систем и компьютерных систем обучения.

В результате освоения дисциплины студент должен:

знать основные принципы разработки и создания лингвистических корпусов и ресурсов; преимущества и недостатки формальных моделей, лежащих в основе различных модулей автоматической обработки текста; необходимые этапы морфологического анализа и проблемы, возникающие при моделировании каждого из этапов; основные алгоритмы, используемые для построения автоматического синтаксического анализа; наиболее известные доступные для свободного использования компоненты автоматического анализа;

уметь проводить оценку качества систем автоматического морфологического, синтаксического и семантического анализа; использовать соответствующие модули в различных приложениях; участвовать в разработке отдельных модулей в системах автоматического анализа текстов.

Процесс изучения дисциплины направлен на формирование следующих компетенций: ОПК1, ОПК4, ПК2, ПК5, ПК7, ПК12, ПК13, ПК17

## **2. МЕСТО ДИСЦИПЛИНЫ В ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЕ**

Дисциплина относится к профессиональному циклу дисциплин по выбору студента вариативной части учебного плана.

Базируется на ранее изученных дисциплинах: «Компьютерная логика», «Программирование», «Иностранный язык».

### 3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

#### 3.1. Распределение учебных часов по темам дисциплины и видам занятий

Наименование тем (содержательных модулей)	Количество часов				
	Всего	В том числе			
		Лекции	Практ. (Семин.)	Лабор.	СРС
Тема 1. Введение. Лингвистика как наука.	6	2	2	-	2
Тема 2. Семиотика. Взаимосвязь с лингвистикой.	6	2	2	-	2
Тема 3. Естественные и искусственные языки. Языки человеко-машинного общения и программирования как искусственные языки.	6	2	-	2	2
Тема 4. Моделирование как основной метод в прикладной и компьютерной лингвистике.	6	2	-	2	2
Тема 5. Информационный поиск.	15	4	2	4	5
Тема 6. Лексикография как одно из важных направлений прикладной лингвистики.	27	4	2	12	9
Тема 7. Машинный перевод.	6	4	-	-	2
Тема 8. Автоматическая обработка текста. Текст и гипертекст.	9	2	-	4	3
Тема 9. Классификация и кластеризация документов.	9	4	2	-	3
Тема 10. Построение алгоритмов лингвистического анализа.	15	4	2	4	5
Тема 11. Автоматическое реферирование	11	2	-	6	3
Тема 12. Перспективные направления лингвистики.	8	2	3	-	3
Курсовое проектирование	38	-	2	-	36
Подготовка к экзамену	54	-	-	-	-
Итого:	216	34	17	34	77

#### 3.2. Лекции

Тема 1. Введение. Лингвистика как наука.

##### Содержание темы 1:

Система лингвистических дисциплин и направлений. Различные подходы к определению термина «Прикладная лингвистика». Теоретическая и прикладная (практическая) лингвистика. Задачи и направления прикладной лингвистики. Связь лингвистики с другими науками – естественными и гуманитарными.

Литература к теме 1: [1, 2, 3, 4, 9, 10]

## Тема 2. Семиотика. Взаимосвязь с лингвистикой.

### Содержание темы 2:

Знак и классификации знаков. Язык как знаковая система. Информация. Виды и представление информации. Получение, хранение, обработка и передача вербальной информации. Информатика как наука. Взаимосвязь с лингвистикой.

Литература к теме 2: [1, 3, 5, 9, 10]

## Тема 3. Естественные и искусственные языки. Языки человеко-машинного общения и программирования как искусственные языки.

### Содержание темы 3:

Формальные методы описания искусственных языков. Понятие формальной грамматики и языка. Понятие метаязыка. БНФ-нотации и синтаксические диаграммы. Математическая лингвистика и теория формальных языков.

Литература к теме 3: [2, 5, 11-15]

## Тема 4. Моделирование как основной метод в прикладной и компьютерной лингвистике.

### Содержание темы 4:

Понятия «модель» и «лингвистическая модель». Модели знаний и семантики в прикладной лингвистике и искусственном интеллекте.

Языки представления знаний как вариант искусственных языков. Их классификация.

Литература к теме 4: [2, 5, 11-15]

## Тема 5. Информационный поиск.

### Содержание темы 5:

Виды поиска, характеристики информационного поиска. Лингвистика в задачах информационного поиска.

Информационно-поисковые языки как искусственные языки. Классификация информационно-поисковых языков.

Литература к теме 5: [2, 5, 9-15]

## Тема 6. Лексикография как одно из важных направлений прикладной лингвистики.

### Содержание темы 6:

Словарь как объект лексикографии. Классификация и организация словарей. Традиционная и машинная лексикография. Основные направления компьютерной лексикографии. Некоторые особенности автоматических словарей.

Идеографические словари и тезаурусы. Национальный корпус русского языка. Терминография. Терминологические банки данных.

Литература к теме 6: [1-3, 5, 9, 10, 16]

## Тема 7. *Машинный перевод.*

### Содержание темы 7:

Задача машинного перевода как одна из важнейших задач прикладной лингвистики. Хронология развития. Виды. Подходы к моделированию машинного перевода (прямой, через интерлингву, через трансфер).

Технологии и основные этапы работы системы машинного перевода. Соотношение анализа и синтеза текста. Понятие трансфера.

Литература к теме 7: [1-3, 9-15]

## Тема 8. *Автоматическая обработка текста. Текст и гипертекст.*

### Содержание темы 8:

Автоматическая обработка текста. Текст и гипертекст. Компьютерная лингводидактика.

Литература к теме 8: [1-3, 9-15]

## Тема 9. *Классификация и кластеризация документов.*

### Содержание темы 9:

Классификации документов, типы классификаций. Законы Ципфа и Хипса. Модель TF\*IDF.

Кластеризация. Контент-анализ. Извлечение фактов и знаний.

Литература к теме 9: [1, 5, 17]

## Тема 10. *Построение алгоритмов лингвистического анализа.*

### Содержание темы 10:

Разработка алгоритмов анализа естественно-языковых текстов на фонемном, морфемном и морфологическом уровнях.

Разработка алгоритмов анализа естественно-языковых текстов на синтаксическом и прагматическом уровнях.

Литература к теме 10: [2, 5, 11-15]

## Тема 11. *Автоматическое реферирование.*

### Содержание темы 11:

Создание программ сжатого представления естественно-языковых текстов.

Литература к теме 11: [1-3, 5]

## Тема 12. *Перспективные направления лингвистики*

### Содержание темы 12:

Корпусная лингвистика, политическая лингвистика, когнитивная, медиалингвистика и др.

Литература к теме 12: [1-4, 9-15]

### 3.3. Практические (семинарские) занятия

№ п/п	Тема занятия	Объем, час.	Литература
1	<i>Понятие о языке как средстве коммуникации. Речь и язык – предмет и объект языкознания. Структурные уровни языка. Языковые системы.</i>	2	[1, 7,]
2	<i>Семантический треугольник. Информация – понимание в разных науках.</i>	2	[1, 7]
3	<i>Выдача технического задания по курсовому проекту</i>	2	[8]
4	<i>Виды поиска, характеристики информационного поиска.</i>	2	[1, 2, 7]
5	<i>Классификация и организация словарей.</i>	2	[1, 7]
6	<i>Классификации документов. Законы Ципфа и Хинса. Модель TF*IDF.</i>	2	[1, 7]
7	<i>Понятие о фонемном, морфемном, морфологическом синтаксическом и прагматическом уровнях анализа естественно-языковых текстов.</i>	2	[7]
8	<i>Перспективные направления компьютерной лингвистики.</i>	3	[9-15]
Итого:		17	

### 3.4. Лабораторные работы

№ п/п	Тема работы	Объем, час.	Литература
1	<i>Знакомство с компьютерными лингвистическими программами.</i>	4	[6]
2	<i>Знакомство с национальным корпусом русского языка.</i>	2	[6, 16]
3	<i>Типы компьютерных словарей и работа с ними.</i>	4	[6, 9, 10]
4	<i>Разработка компьютерного терминологического словаря для заданной предметной области.</i>	6	[6, 9, 10]
5	<i>Поиск текстовой описательной информации в тексте программы на языке С.</i>	4	[6]
6	<i>Поиск информации для подготовки прототипов функций в программе.</i>	4	[6]
7	<i>Разработка программы автоматической подготовки документации к программе на языке С.</i>	10	[6, 9, 10]
Итого:		34	

### 3.5. Самостоятельная работа студента

№ п/п	Виды самостоятельной работы студента	Объем, час.
1	Изучение лекционного материала (не менее 50% от объема лекций)	17
2	Подготовка к практическим занятиям (не менее 50% от объема аудиторных практических занятий)	7
3	Подготовка к лабораторным работам (не менее 50% от объема аудиторных лабораторных занятий)	17
4	Выполнение курсового проекта (36 часов)	36
5	Выполнение курсовой работы (27 часов)	-
6	Выполнение индивидуального задания (не менее 9 часов)	-
Итого:		77

### 3.6. Курсовой проект (работа), индивидуальное задание

В ходе выполнения курсового проекта студент получает профессиональные навыки в анализе технического задания, включая постановку задачи, анализ требований к модели, выбор форм и форматов представления исходных данных и результатов.

Проектирование лингвистической системы в соответствии с индивидуальным заданием предусматривает: анализ входных и выходных данных; разработку спецификаций; выбор/разработку алгоритма; обоснование выбора программных средств реализации/моделирования алгоритма; создание, тестирование и отладку модели; оформлении документации, пояснительной записки и графической части в соответствии с [7, 8].

Объем учебной нагрузки при выполнении курсового проекта – 36 часов.

Рекомендуемый объем пояснительной записки по курсовому проекту – не более 60 страниц формата А4 (210×297 мм), включая фрагменты листинга программного кода. Объем графической части – 2 листа формата А1.

## 4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

**Текущий контроль** знаний студентов производится во время контрольных опросов в ходе проведения лабораторных и практических занятий.

**Промежуточная аттестация** по результатам освоения дисциплины в семестре проводится в форме защиты курсового проекта и сдачи экзамена в соответствии с «Положением об организации учебного процесса в Донецком национальном техническом университете (новая редакция)», утвержденном 25.11.2016 года, протокол №8.

Для определения уровня знаний студентов преподаватель руководствуется критериями оценки знаний, являющимися составляющей учебно-методического комплекса дисциплины.

## 5. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

### Литература:

#### Основная:

1. Бугаева В. Ю. *Иностранный язык профессиональной направленности (английский язык) = Бугаєва В. Ю. Іноземна мова професійного спрямування (англійська мова) : сб. тезисов и заданий / В. Ю. Бугаева. – Харьков, 2010. – 267 с.*

2. Вовчасная Н. Я. *Профессиональный немецкий язык. Тесты и упражнения = Вовчасна Н.Я. Професійна німецька мова. Тести і вправи / Н. Я. Вовчасная. – Львов, 2013. – 213 с*



3. Попова И. Н. Французский язык : учеб. для 1 курса вузов / И. Н. Попова. – Харьков : Нестор Академик Паблшер, 2010. – 576 с.

Дополнительная:

4. Учебное пособие для проведения практических занятий по английскому языку научных и технических профессий для направления подготовки "Электротехника и электротехнологии" [Электронный ресурс]. Ч. 1 / А. В. Корощенко [и др.] ; ГВУЗ "ДонНТУ", Каф. электромеханики и ТОЭ, Каф. англ. яз. - (2,5 Мб). - Донецк : ГВУЗ "ДонНТУ", 2011. - 1 файл. - Систем. требования: ZIP-архиватор, Microsoft Word.

5. Попова И. Н. Французский язык [Электронный ресурс] = Manuel de Français : учебник для 1 курса вузов и факультетов иностранных языков / И. Н. Попова, Ж. А. Казакова, Г. М. Ковальчук. - Изд. 21-е, испр. - 75 Мб. - Москва : Нестор Академик Паблшерз, 2009. - 1 файл. - Систем. требования: Acrobat Reader.

**Учебно-методические издания, разработанные в ДонНТУ:**

К самостоятельной работе студента:

6. Методические рекомендации относительно организации самостоятельной работы студентов по дисциплине «Компьютерная лингвистика»

К практическим занятиям:

7. Методические указания к выполнению практических занятий по курсу «Компьютерная лингвистика» .

**Интернет-ресурсы**

8. Сведения и основы = Data&Basement, <http://www.rbkc.gov.uk/wamdocs/Basements%20Data%20Report%20v2.pdf>

9. Экономист = The Economist, [https://www.researchgate.net/publication/258120737\\_PDF\\_The\\_economist](https://www.researchgate.net/publication/258120737_PDF_The_economist)

Международный журнал механики = INTERNATIONAL JOURNAL OF MECHANICAL SCIENCES, <https://www.elsevier.com/journals/international-journal-of-mechanical-sciences/0020-7403?generatepdf=true>

10. Компьютерный и электротехнический инжиниринг = COMPUTERS & ELECTRICAL ENGINEERING, <https://www.elsevier.com/journals/computers-and-electrical-engineering/0045-7906?generatepdf=true>

11. Журнал исследований в области компьютерного и электротехнического инжиниринга = Journal of Electrical and Electronics Engineering Research, режим доступа: [http://www.academicjournals.org/ebook/journal1402474084\\_June%202014%20Issue.pdf](http://www.academicjournals.org/ebook/journal1402474084_June%202014%20Issue.pdf)


## **6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **1. Лекционные занятия:**

- аудитория № 4.035, оснащенная презентационной техникой: проектор, экран;
- комплект электронных презентаций.

### **2. Лабораторные работы:**

- лаборатория (к.4.19), оснащенная 9 ПК.

Составитель рабочей программы:  Мальчева Р.В.  
(подпись)