



Авторы-составители:

**Семина Татьяна Алексеевна**, к.ф.н., ассистент кафедры теоретической и прикладной лингвистики.

Рабочая программа дисциплины «Основы компьютерной лингвистики» составлена в соответствии с требованиями Федерального Государственного образовательного стандарта высшего образования по направлению подготовки 45.03.02 Лингвистика, утверждённом приказом МИНОБРНАУКИ РОССИИ от 12. 08. 2020 года №969.

Дисциплина относится к части, формируемой участниками образовательных отношений блока 1, и является обязательной дисциплиной.

Год начала подготовки 2021

## СОДЕРЖАНИЕ

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ.....	4
1.1. Цель и задачи дисциплины.....	4
1.2. Планируемые результаты обучения .....	4
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ....	4
3. ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	4
3.1. Объем дисциплины .....	4
3.2. Содержание дисциплины.....	5
4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ.....	6
5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ.....	7
5.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы .....	7
5.2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания.....	7
5.3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы .....	11
5.3.1. Примерные вопросы и задания для текущего контроля:.....	11
5.3.2. Примерный список рекомендованных тем докладов .....	12
5.3.3. Примерные темы индивидуальных и групповых проектов/сообщений .....	13
5.3.4. Промежуточная аттестация .....	13
5.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.....	15
5.4.1. Этап: проведение текущего контроля успеваемости по дисциплине «Основы компьютерной лингвистики» .....	15
5.4.2. Промежуточная аттестация .....	16
6. УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	18
6.1 Основная литература: .....	18
6.2 Дополнительная литература: .....	18
6.3. Ресурсы информационно-телекоммуникационной сети «Интернет» .....	19
6. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ .....	<b>Error!</b>
<b>Bookmark not defined.</b>	
8. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ ...	<b>Error! Bookmark not defined.</b>
9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ ....	<b>Error!</b>
<b>Bookmark not defined.</b>	



## 1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

### 1.1. Цель и задачи дисциплины

Целью программы «Основы компьютерной лингвистики» является формирование у студентов представлений о множестве форматов текстовых документов, навыков обработки русскоязычных и иноязычных текстов в производственно-практических целях, разработки средств информационной поддержки лингвистических областей знаний.

### 1.2. Планируемые результаты обучения

В результате освоения данной дисциплины у обучающихся формируются следующие компетенции:

СПК-1 - способен использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач

СПК-2 - Способен применять систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях о закономерностях функционирования языков мира

СПК-3- Способен применять основные современные методы научного исследования, в том числе и в смежных областях, в самостоятельных исследованиях

СПК-4 - Владеет основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических данных

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к части, формируемой участниками образовательных отношений блока 1, и является обязательной к изучению.

В методическом плане дисциплина опирается на знания, полученные в средней школе. Данная учебная дисциплина дает начальную теоретическую базу, необходимую для формирования общекультурного кругозора студентов. Это теоретический фундамент, на котором основываются все языковые и переводческие дисциплины в формировании у студентов базовых лингвистических знаний, таких как теоретическая фонетика, теоретическая грамматика и лексикология, позволяет установить с ними меж предметные связи.

## 3. ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

### 3.1. Объем дисциплины

Показатель объема дисциплины	Форма обучения
	Очная
Объем дисциплины в зачетных единицах	6
Объем дисциплины в часах	216
Контактная работа:	86.7
Лекции	14
Практические занятия	72
Контактные часы на промежуточную аттестацию	0,4
Зачет	0.4

Курсовая работа	0.3
Самостоятельная работа	114
Контроль	33.3

Формой промежуточной аттестации является зачет с оценкой в 3 и 4 семестрах и курсовая работа в 4 семестре.

### 3.2. Содержание дисциплины

Наименование разделов (тем) Дисциплины с кратким содержанием	Виды занятий	
	Лекции	Практические занятия
<b>Тема 1. Компьютерная лингвистика: цели, задачи и методы. Современные приложения компьютерной лингвистики.</b> Определение компьютерной лингвистики, связь с другими дисциплинами. История развития компьютерной лингвистики. Библиотеки инструментов для обработки текста.	10	-
<b>Тема 2. Первые этапы обработки текста.</b> Необходимость обработки «сырого» текста. Токенизация. Другие этапы предобработки текста.	-	4
<b>Тема 3. Регулярные выражения.</b> Понятие регулярных выражений. Метасимволы регулярных выражений. Применение регулярных выражений для обработки языка.	-	8
<b>Тема 4. Компьютерная морфология.</b> Основные понятия. Стемминг и лемматизация: алгоритмы и различия.		8
<b>Тема 5. Конечно-автоматная морфология.</b> Конечный автомат как способ организации словаря для морфологического анализа и синтаксиса.		8
<b>Тема 6. Скрытые Марковские модели.</b> Основная структура. Применение для обработки морфологии.		8
<b>Тема 7. Задание правил для анализа морфологии.</b> Типы правил для обработки фонологических особенностей. Создание набора правил для обработки морфологии. Теория оптимальности.		8
<b>Тема 8. Особенности создания лингвистических процессоров и обработки семантики и прагматики.</b> Пирамида обработки естественного языка.	4	-
<b>Тема 9. Особенности обработки синтаксиса.</b>		8

Порождающие грамматики. Исчисление Ламбека. SKY-парсинг.		
<b>Тема 10. Представление текста и слова в виде вектора значений.</b> Задание документа в виде вектора: способы, применение векторной модели. Дистрибутивная гипотеза. Word2vec.		8
<b>Тема 11. Лямбда-исчисление.</b> Задание значения предложения в виде предикатов.		12
Итого	14	72

Формой промежуточной аттестации является зачет в 3 и 4 семестрах и курсовая работа в 4 семестре.

#### 4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Темы для самостоятельного изучения	Изучаемые вопросы	Количество часов	Формы самостоятельной работы	Методические обеспечения	Формы отчетности
Основы обработки текста	Предобработка, регулярные выражения	16	Чтение теории по обсуждаемым вопросам	Рекомендуемая литература интернет-ресурсы записанное преподавателем виде	Практическое задание Тест
Библиотеки инструментов для анализа данных.	1. Natural Language ToolKit 2. Anaconda 3. Доступные парсеры и пост-теггеры: обзор и анализ системы.	16	Чтение теории по обсуждаемым вопросам	Учебно-методическая литература интернет-ресурсы	Практическое задание
Лемматизация и стемминг	Алгоритмы лемматизации и стемминга Написание алгоритма стемминга для русского языка	18	Чтение теории по обсуждаемым вопросам	Учебно-методическая литература интернет-ресурсы	Практическое задание, тест
Лексические преобразователи	Конечные автоматы Написание автомата для фрагмента морфологии русского языка	16	Чтение теории по обсуждаемым вопросам	Учебно-методическая литература интернет-ресурсы	Практическое задание
Формальные языки и грамматики	Формальное представление грамматики русского языка	16	Чтение теории по обсуждаемым вопросам	Учебно-методическая литература интернет-	Практическое задание

				ресурсы	
Векторная модель	Представление документа в виде вектора Представление значения слова в виде вектора	16	Чтение теории по обсуждаемым вопросам	Учебно-методическая литература интернет-ресурсы	Практическое задание
Формальная семантика	Ассерция и презумпция Построение семантического дерева	16	Чтение теории по обсуждаемым вопросам	Учебно-методическая литература интернет-ресурсы	Практическое задание
<b>Итого</b>		114			

## **5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ**

### **5.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы**

<b>Код и наименование компетенции</b>	<b>Этапы формирования</b>
<b>СПК-1</b> - способен использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач	1.Работа на учебных занятиях 2.Самостоятельная работа
<b>СПК-2</b> - Способен применять систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях о закономерностях функционирования языков мира	1.Работа на учебных занятиях 2.Самостоятельная работа
<b>СПК-3</b> - Способен применять основные современные методы научного исследования, в том числе и в смежных областях, в самостоятельных исследованиях	1.Работа на учебных занятиях 2.Самостоятельная работа
<b>СПК-4</b> - Владеет основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических данных	1.Работа на учебных занятиях 2.Самостоятельная работа

### **5.2 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания**

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
-------------------------	--------------------------	-------------------	----------------------	---------------------	------------------

СПК-1	Пороговый	1.Работа на учебных занятиях 2.Самостоятельная работа	<b>Знать:</b> основные понятия теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации <b>Уметь:</b> применять полученные знания для решения профессиональных задач	Текущий контроль: Посещение занятий Выступление с докладом сообщением Тесты Задания для самостоятельного изучения	41–60 баллов
	Продвинутый	1.Работа на учебных занятиях 2.Самостоятельная работа	<b>Знать:</b> основные понятия теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации <b>Уметь:</b> применять полученные знания для решения профессиональных задач <b>Владеть:</b> Системным подходом при решении поставленных задач в избранной профессиональной сфере.	Текущий контроль: Посещение занятий Выступление с докладом / сообщением Участие в групповом или индивидуальном проекте Тесты Задания для самостоятельного изучения	61–100 баллов
СПК-2	Пороговый	1.Работа на учебных занятиях 2.Самостоятельная работа	<b>Знать:</b> Основные фонетические, лексические, грамматические, словообразовательные явления в языках мира; системы синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации <b>Уметь:</b> Применять полученные знания о структуре и системе языков мира в самостоятельных исследованиях	Текущий контроль: Посещение занятий Выступление с докладом сообщением Тесты Задания для самостоятельного изучения	41–60 баллов
	Продвинутый	1.Работа на учебных занятиях 2.Самостоятельная работа	<b>Знать:</b> Основные фонетические, лексические, грамматические, словообразовательные явления в языках мира;	Текущий контроль: Посещение занятий Выступление с докладом /	61–100 баллов

			<p>системы синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации</p> <p><b>Уметь:</b> Применять полученные знания о структуре и системе языков мира в самостоятельных исследованиях</p> <p><b>Владеть:</b> Комплексным подходом к решению лингвистических задач на разных уровнях строения языка с учетом специфики его функционирования; навыками работы с программными средствами и информационными ресурсами филологической направленности, методикой работы с лингвистическими базами данных</p>	<p>сообщением</p> <p>Участие в групповом или индивидуальном проекте</p> <p>Тесты</p> <p>Задания для самостоятельного изучения</p>	
СПК-3	Пороговый	<p>1.Работа на учебных занятиях</p> <p>2.Самостоятельная работа</p>	<p><b>Знать:</b> общенаучные методы и конкретные методики изучения данных в соответствующей области лингвистики; принципы работы с библиографическими источниками</p> <p><b>Уметь:</b> использовать основные информационно-поисковые и экспертные системы, системы представления знаний в данной предметной области, принципы научно-доказательного изложения материала.</p>	<p>Текущий контроль:</p> <p>Посещение занятий</p> <p>Выступление с докладом</p> <p>сообщением</p> <p>Тесты</p> <p>Задания для самостоятельного изучения</p>	41–60 баллов
	Продвинутый	<p>1.Работа на учебных занятиях</p> <p>2.Самостоятельная работа</p>	<p><b>Знать:</b> общенаучные методы и конкретные методики изучения данных в соответствующей области лингвистики; принципы работы с библиографическими источниками</p> <p><b>Уметь:</b> использовать основные</p>	<p>Текущий контроль:</p> <p>Посещение занятий</p> <p>Выступление с докладом / сообщением</p> <p>Участие в групповом или индивидуальном проекте</p>	61–100 баллов

			<p>информационно-поисковые и экспертные системы, системы представления знаний в данной предметной области, принципы научно-доказательного изложения материала.</p> <p><b>Владеть:</b> проблематикой смежных с лингвистикой областей и возможными подходами к их решению с позиций комплексного подхода</p>	<p>Тесты Задания для самостоятельного изучения</p>	
СПК-4	Пороговый	<p>1.Работа на учебных занятиях 2.Самостоятельная работа</p>	<p><b>Знать:</b> математико-статистические методы обработки лингвистической информации, основы программирования, принципы автоматической обработки корпусов текстов</p> <p><b>Уметь:</b> применять полученные знания для анализа и обработки нового лингвистического материала на изучаемых языках</p>	<p>Текущий контроль: Посещение занятий Выступление с докладом сообщением Тесты Задания для самостоятельного изучения</p>	41–60 баллов
	Продвинутый	<p>1.Работа на учебных занятиях 2.Самостоятельная работа</p>	<p><b>Знать:</b> математико-статистические методы обработки лингвистической информации, основы программирования, принципы автоматической обработки корпусов текстов</p> <p><b>Уметь:</b> применять полученные знания для анализа и обработки нового лингвистического материала на изучаемых языках</p> <p><b>Владеть:</b> способами представления полученных результатов, методикой изложения, принятой в соответствующей области лингвистического знания</p>	<p>Текущий контроль: Посещение занятий Выступление с докладом / сообщением Участие в групповом или индивидуальном проекте Тесты Задания для самостоятельного изучения</p>	61–100 баллов

### 5.3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

#### 5.3.1. Примерные вопросы и задания для текущего контроля:

1. Постройте значение предложения при помощи вывода из синтаксического дерева. Дан результат автоматического парсинга, используйте его для построения дерева НС (правильность именованя вершин не будет учитываться). Добавьте необходимые функциональные вершины.

```
(ROOT
(S
(NP (PRP I))
(VP (VBP want)
(S
(NP (DT the) (NN dog))
(VP (TO to)
(VP (VB eat)
(NP (PRP$ my) (NN homework))))))
(. .)))
```

I want the dog to eat my homework.

Посчитайте расстояние Левенштейна для пары слов:

Распределение – уравнение

2. Назовите основные этапы предварительной обработки текста.

3. Объясните необходимость перевода грамматики в нормальную форму Хомского при парсинге.

3. Раскройте понятие «чанкинг».

4. Опишите преимущества IOB разметки при проведении чанкинга.

5. Назовите различия между двумя подходами к нормализации словоформы: стеммингом и лемматизацией.

6. Объясните, какой способ описания алгоритма больше всего подходит для описания лингвистической системы.

7. Дан оригинальный текст и этот текст после обработки. Опишите, какие изменения произошли и задайте набор правил, по которым производилось преобразование:

Оригинал:

Stoker was born on 8 November 1847 at 15 Marino Crescent, Clontarf, on the northside of Dublin, Ireland.[1] His parents were Abraham Stoker (1799–1876) from Dublin and Charlotte Mathilda Blake Thornley (1818–1901), who was raised in County Sligo.[2] Stoker was the third of seven children, the eldest of whom was Sir Thornley Stoker, 1st Bt.

Текст после стемминга (при помощи Porter Stemmer):

Stoker wa born on 8 Novemb 1847 at 15 Marino Crescent , Clontarf , on the northsid of Dublin , Ireland . [ 1 ] Hi parent were Abraham Stoker ( 1799 – 1876 ) from Dublin and Charlott Mathilda Blake Thornley ( 1818 – 1901 ) , who wa rais in Counti Sligo . [ 2 ] Stoker wa the third of seven children , the eldest of whom wa Sir Thornley Stoker , 1st Bt .

#### Примерные тестовые вопросы:

1. Что из перечисленного является свойствами алгоритмов?

1. понятность, адекватность, корректность, конечность
2. детерминированность, результативность, конечность, понятность
3. дискретность, объективность, понятность, конечность
4. детерминированность, универсальность, объективность, понятность

Ответ: 2

2. Какой из компонентов НЕ относится к предварительной обработке твитов:
1. замена ссылок на общее условное обозначение
  2. удаление эмодиконов
  3. замена всех букв на строчные
  4. токенизация

Ответ: 2

3. Какой из алгоритмов подойдет для токенизации китайского текста:
1. MaxMatch
  2. WordPunctTokenizer
  3. WhitespaceTokenizer
  4. Byte-pair encoding

Ответ: 1

4. Какие из представленных задач относятся к полноценным системам анализа данных:
1. Извлечение и классификация именованных сущностей
  2. Реферирование текста
  3. Рубрикация текстов
  4. Частичный парсинг

Ответ: 2 и 3

### **5.3.2. Примерный список рекомендованных тем докладов**

1. Сравнение словесного и графического способов описания алгоритма.
2. Масштабирование данных с помощью tf-idf.
3. Алгоритм Витерби в лингвистических задачах.
4. Применение конечных автоматов в лингвистических задачах.
5. Применение регулярных выражений для извлечения данных
6. Принципы создания контролируемого языка
7. Описание морфологических правил русского языка при помощи конечных преобразователей
8. Викификация как современный подход к созданию лингвистических ресурсов
9. Особенности морфологического анализа русского языка.
10. Применение динамического программирования для автоматического синтаксического анализа.
11. Типология формальных языков Ноама Хомского.
12. Перевод правил контекстно-свободной грамматики в нормальную форму Хомского.
13. Синтаксис языков программирования.
14. Частеречная разметка как основной компонент морфологического анализа.
15. Синтаксическое исчисление Ламбека.
16. Лямбда-исчисление.
17. Кластеризация данных из социальных сетей.
18. Методы сбора данных для создания лингвистических баз данных и корпусов для машинного обучения.
19. Характеристика «жадных» алгоритмов
20. Частичное машинное обучение с учителем

### **5.3.3. Примерные темы индивидуальных и групповых проектов/сообщений**

1. Составить алгоритм токенизации.
2. Сравнение подходов к описанию алгоритма у разных исследователей.
3. Формальное описание алгоритма извлечения данных их неструктурированных текстов.
4. Сбор и разметка данных для системы анализа тональности.
5. Сравнение доступных в сети библиотек с инструментами для анализа текста.
6. Описание набора фонологических правил для морфологического анализатора при помощи формализма Хомского-Халле и при помощи теории оптимальности.

### **5.3.4. Промежуточная аттестация**

Промежуточная аттестация осуществляется в форме зачета с оценкой в 3 и 4 семестрах и курсовой работы в 4 семестре.

#### **Примерные вопросы к зачету**

1. Конечные автоматы: моделирование словоизменительной и словообразовательной морфологии.
2. Нормализация текста, этапы
3. Токенизация
4. Лемматизация и стемминг
5. Подходы к анализу морфологии
6. Частеречная разметка при помощи Скрытых Марковских моделей
7. Регулярные выражения
8. Расстояние редактирования (Edit distance)
9. Дистрибутивная гипотеза в семантике
10. Семантический вектор
11. Проблема двусмысленности при автоматическом синтаксическом анализе
12. SKY-парсинг
13. Вероятностные контекстно-свободные грамматики: преимущества при парсинге
14. Лексикализованный парсинг
15. Парсинг зависимостей и проект Universal dependencies
16. Типы словосочетаний, извлечение словосочетаний
17. Лямбда-исчисление: лямбды, предикаты, кванторы, теория миров.
18. Ассерция и презумпция.

#### **Примерные темы курсовых работ:**

1. Алгоритм k-ближайших соседей
2. Алгоритм Витерби
3. Кодировка текста
4. Алгоритмы поиска плагиата
5. Атрибуция текста
6. Векторное представление слов
7. Викификация как современный подход к созданию лингвистических ресурсов
8. Генетические алгоритмы
9. Деревья решений

10. Дискретно-геометрический подход к распознаванию символов.
11. Дистрибутивная семантика
12. Идентификация языка
13. Иерархическая кластеризация
14. Извлечение терминов
15. Инструменты для анализа текста в современных текстовых редакторах
16. Использование данных из социальных сетей для лингвистических программ
17. Исчисление Ламбека
18. Классификация формальных языков Н. Хомского
19. Кластеризация при помощи алгоритма k-средних
20. Компьютерная психолингвистика: основные направления исследований
21. Корпус текстов как материал для машинного обучения
22. Корпусы речевых данных
23. Логистическая регрессия
24. Лямбда-исчисление в лингвистических задачах.
25. Масштабирование данных с помощью tf-idf
26. Метод опорных векторов
27. Методы выделения именных групп
28. Методы разрешения анафоры
29. Методы частеречной разметки для русского языка
30. Мультимодальные корпуса
31. Наивный Байесовский классификатор
32. Нейронные сети прямого распространения
33. Объектная модель документа
34. Онтологический подход к семантике
35. Описание морфологических правил русского языка при помощи конечных преобразователей
36. Основные подходы к автоматическому синтаксическому анализу
37. Подходы к дизамбигуации
38. Подходы к извлечению именованных сущностей
39. Предобработка текста
40. Применение конечных автоматов в лингвистических задачах
41. Применение регулярных выражений для извлечения данных
42. Принципы создания контролируемого языка
43. Псевдокод как инструмент описания алгоритмов
44. Рекуррентные нейронные сети
45. Самоорганизующиеся карты Кохонена
46. Сверточные нейронные сети
47. Сравнительный анализ информационно-поисковых систем
48. Сравнительный анализ морфологических анализаторов для английского языка
49. Сравнительный анализ морфологических анализаторов для русского языка
50. Сравнительный анализ систем распознавания речи
51. Сравнительный анализ систем синтаксического анализа
52. Сравнительный анализ систем синтеза речи
53. Структура гипертекста
54. Тезаурусы русского языка
55. Теория оптимальности для описания правил
56. Токенизация для языков с иероглифическим письмом
57. Характеристики «жадных» алгоритмов
58. Частичное обучение с учителем

#### 5.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Промежуточная аттестация по дисциплине «Основы компьютерной лингвистики» учитывает уровень результатов обучения, общее качество работы студента, дисциплинированность, самостоятельность. Освоение дисциплины оценивается по балльной шкале.

Общее количество баллов по дисциплине – 100 баллов.

Максимальное количество баллов, которое может набрать студент в течение семестра за посещаемость, выполнений заданий для самостоятельного изучения, участие в групповом или индивидуальном проекте, прохождение тестов, выступление с докладом/сообщением равняется 60 баллам.

Максимальная сумма баллов, которые студент может набрать на зачете, равняется 40 баллам.

##### 5.4.1. Этап: проведение текущего контроля успеваемости по дисциплине «Основы компьютерной лингвистики»

Вид работы	Пятибалльная шкала	Шкала оценивания
1. Посещение занятий (лекции, практические занятия)	«отлично»	<b>10 баллов</b> , если студент посетил 90% от всех занятий
	«хорошо»	<b>8 баллов</b> , если студент посетил как минимум 70% от всех занятий
	«удовлетворительно»	<b>5 баллов</b> если студент посетил как минимум 50% от всех занятий
	«неудовлетворительно»	<b>3 балла</b> , если из всех занятий студент посетил как минимум 25%
2. Ответы на практическом занятии	«отлично»	<b>30 баллов</b> ответ полный, с привлечением знаний из разных разделов курса; 5 и более примеров
	«хорошо»	<b>20 баллов</b> ответ полный; 3-4 примера
	«удовлетворительно»	<b>15 баллов</b> ответ неполный; 1-2 примера
	«неудовлетворительно»	<b>13 баллов</b> ответ не соответствует теоретическому вопросу; без примеров
3. Тестирование	«отлично»	<b>25 баллов</b> , если из всех заданий студент выполнил как минимум 70%

	«хорошо»	<b>22 баллов</b> , если из всех заданий студент выполнил как минимум 60%
	«удовлетворительно»	<b>17 баллов</b> , если из всех заданий студент выполнил 40%-59%
	«неудовлетворительно»	<b>9 баллов</b> , если из всех заданий студент выполнил 20%
4. Самостоятельная подготовка к практическим занятиям	«отлично»	<b>25 баллов</b> , если из всех заданий студент выполнил как минимум 80%
	«хорошо»	<b>22 баллов</b> , если из всех заданий студент выполнил как минимум 60%
	«удовлетворительно»	<b>17 баллов</b> , если из всех заданий студент выполнил 40%
	«неудовлетворительно»	<b>11 баллов</b> , если из всех заданий студент выполнил 20%
5. Зачет с оценкой	«отлично»	<b>10 баллов</b>
	«хорошо»	<b>8 баллов</b>
	«удовлетворительно»	<b>6 баллов</b>
	«неудовлетворительно»	<b>4 балла</b>

### 5.4.2. Промежуточная аттестация

#### Критерии оценки на зачете

При оценке знаний на зачете учитываются:

1. Понимание и степень усвоения теории курса.
2. Уровень знания фактического материала в объёме программы.
3. Правильность формулировки основных понятий и закономерностей.
4. Логика, структура и грамотность изложения вопроса.
5. Использование примеров из монографической литературы (статьи, хрестоматии, художественные произведения).
6. Умение связать теорию с практическим применением.
7. Умение сделать обобщение, выводы.
8. Умение ответить на дополнительные вопросы.
9. Глубокое и прочное усвоение знаний программного материала (умение выделять главное, существенное).
10. Знание авторов-исследователей по данной проблеме.

#### Шкала оценивания ответа на зачете

Уровень овладения	неудовлетворительный	удовлетворительный	оптимальный	высокий
Дескрипторы				
Полнота ответа на теоретический вопрос	4 Ответ, не соответствующий теоретическому вопросу	5 Ответ неполный	7 Ответ полный	8 Ответ полный, с привлечением знаний из разных разделов курса
Знание терминологии, умение давать определения понятиям	4 отсутствует	5 Определения даются с некоторыми неточностями	7 Определения даются без собственных объяснений и дополнений	8 Четкие определения, умение объяснить их и дополнить
Знание персоналий, сопряженных с теоретическим вопросом	4 отсутствует	5 1-2 примера	6 3-4 примера	8 5 и более примеров
Умение проиллюстрировать явление практическими примерами	3 отсутствие примеров	5 1-2 примера	6 3-4 примера	8 5 и более примеров
Ответы на вопросы экзаменатора	3 Нет ответов на вопросы	5 Только ответы на элементарные вопросы	6 Ответы на вопросы полные или частично полные	8 Ответы на вопросы полные с приведением примеров и/или пояснений
Итоговый балл (максимальн)	<b>18</b>	<b>25</b>	<b>32</b>	<b>40</b>
	<b>не зачтено</b>	<b>зачтено</b>	<b>зачтено</b>	<b>зачтено</b>

**Сводная шкала оценивания по дисциплине «Основы компьютерной лингвистики»**

Вид работы	«отлично»	«хорошо»	«удовлетворительно»	«неудовлетворительно»
Посещение занятий	5	4	3	2
Выступление с докладом/сообщением	5	4	3	2
Участие в групповом или индивидуальном проекте	20	15	13	11
Прохождение теста	20	15	12	10
Выполнение заданий для самостоятельного изучения	10	7	6	5
Зачет с оценкой	40	35	28	20
<b>Итого</b>	<b>81- 100</b>	<b>61-80</b>	<b>41-60</b>	<b>0-40</b>

- оценка «зачтено» ставится студенту, набравшему 41-100 баллов;

- оценка «не зачтено» ставится студенту, набравшему 0-40 баллов.

## **6. УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **6.1 Основная литература:**

1. Гусякова, А. В. Информационные технологии и лингвистика XXI века : учебное пособие / А. В. Гусякова. – Москва : Московский педагогический государственный университет (МПГУ), 2016. – 96 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=469675> (дата обращения: 11.06.2021). – Библиогр. в кн. – ISBN 978-5-4263-0398-0. – Текст : электронный.
2. Калугян, К. Х. Информатика. Информационные технологии и системы : учебное пособие : [16+] / К. Х. Калугян. – Ростов-на-Дону : Издательско-полиграфический комплекс РГЭУ (РИНХ), 2018. – 80 с. : схем., табл., ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=567017> (дата обращения: 11.06.2021). – Библиогр. в кн. – ISBN 978-5-7972-2466-2. – Текст : электронный.
3. Моисеева, И. Ю. Квантитативная лингвистика и новые информационные технологии : учебное пособие / И. Ю. Моисеева ; Оренбургский государственный университет. – Оренбург : Оренбургский государственный университет, 2017. – 103 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=481797> (дата обращения: 11.06.2021). – Библиогр. в кн. – ISBN 978-5-7410-1713-5. – Текст : электронный.
4. Царёв, Р. Ю. Алгоритмы и структуры данных (CDIO) : учебник / Р. Ю. Царёв, А. В. Прокопенко ; Сибирский федеральный университет. – Красноярск : Сибирский федеральный университет (СФУ), 2016. – 204 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=497016> (дата обращения: 13.06.2021). – Библиогр. в кн. – ISBN 978-5-7638-3388-1. – Текст : электронный.

### **6.2 Дополнительная литература:**

3. Златопольский, Д. М. Программирование: типовые задачи, алгоритмы, методы : [12+] / Д. М. Златопольский. – 4-е изд. (эл.). – Москва : Лаборатория знаний, 2020. – 226 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=222873> (дата обращения: 13.06.2021). – Библиогр. в кн. – ISBN 978-5-00101-789-9. – Текст : электронный.
4. Изюмов, А. А. Компьютерные технологии в науке и образовании : учебное пособие / А. А. Изюмов, В. П. Коцубинский ; Томский Государственный университет систем управления и радиоэлектроники (ТУСУР). – Томск : Эль Контент, 2012. – 150 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=208648> (дата обращения: 11.06.2021). – ISBN 978-5-4332-0024-1. – Текст : электронный.
5. Информационные технологии в лингвистике: учебное пособие (практикум) : [16+] / сост. Ю. А. Пирвердиева ; Северо-Кавказский федеральный университет. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2019. – 182 с. : схем., табл., ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=596213> (дата обращения: 11.06.2021). – Текст : электронный.
6. Ландовский, В. В. Алгоритмы обработки данных : учебное пособие : [16+] / В. В. Ландовский. – Новосибирск : Новосибирский государственный технический

университет, 2018. – 67 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=574809> (дата обращения: 13.06.2021). – Библиогр. в кн. – ISBN 978-5-7782-3645-5. – Текст : электронный.

7. Математическая логика и теория алгоритмов : учебное пособие / сост. А. Н. Макоха, А. В. Шапошников, В. В. Бережной ; Министерство образования Российской Федерации и др. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2017. – 418 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=467015> (дата обращения: 13.06.2021). – Библиогр. в кн. – Текст : электронный.

8. Непейвода, Н. Н. Прикладная логика : учебное пособие : [16+] / Н. Н. Непейвода. – 3-е изд., существ. перераб. и доп. – Москва ; Берлин : Директ-Медиа, 2019. – 576 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=561272> (дата обращения: 11.06.2021). – ISBN 978-5-4499-0126-2. – DOI 10.23681/561272. – Текст : электронный.

9. Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python : [16+] / В. М. Волкова, М. А. Семенова, Е. С. Четвертакова, С. С. Вожов. – Новосибирск : Новосибирский государственный технический университет, 2017. – 74 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=576496> (дата обращения: 13.06.2021). – Библиогр.: с. 48. – ISBN 978-5-7782-3183-2. – Текст : электронный.

10. Теория и практика машинного перевода : учебное пособие / авт.-сост. Э. В. Пиванова ; Северо-Кавказский федеральный университет. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2014. – 115 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=457763> (дата обращения: 11.06.2021). – Текст : электронный.

1. Шелудько, В. М. Основы программирования на языке высокого уровня Python : учебное пособие : [16+] / В. М. Шелудько. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2017. – 147 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=500056> (дата обращения: 13.06.2021). – Библиогр. в кн. – ISBN 978-5-9275-2649-9. – Текст : электронный.

11. Шелудько, В. М. Язык программирования высокого уровня Python: функции, структуры данных, дополнительные модули : [16+] / В. М. Шелудько. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2017. – 108 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=500060> (дата обращения: 13.06.2021). – Библиогр. в кн. – ISBN 978-5-9275-2648-2. – Текст : электронный.

2. Щипицина, Л.Ю. Информационные технологии в лингвистике [Текст] : учеб.пособие / Л. Ю. Щипицина. - 3-е изд., стереотип. - М. : Флинта, 2017. - 128с. - 150-00 (4).

### **6.3.Ресурсы информационно-телекоммуникационной сети «Интернет»**

1. Грамота.ру – русский язык для всех [Электронный ресурс]: справочно-информационный портал. – URL: <http://www.gramota.ru>
2. Издательство «Лань» [Электронный ресурс]: электронно-библиотечная система. – URL: <http://e.lanbook.com/>
3. Издательство «Юрайт» [Электронный ресурс]: электронно-библиотечная система. – URL: <http://biblio-online.ru>
4. Институт русского языка имени В. В. Виноградова Российской академии наук [Электронный ресурс]: официальный сайт. – URL: <http://www.ruslang.ru>
5. Кругосвет [Электронный ресурс]: Универсальная научно-популярная онлайн энциклопедия. – URL: <http://www.krugosvet.ru>

6. Словопедия [Электронный ресурс] : [сайт онлайн-словарей]. – URL: <http://www.slovopedia.com>
7. ibooks.ru[Электронный ресурс]: электронно-библиотечная система. – URL: <http://ibooks.ru>
8. Slovare.ru [Электронный ресурс]: сайт. – URL: <http://www.slovari.ru>
9. Znaniy.com[Электронный ресурс]: электронно-библиотечная система. – URL: <http://znaniy.com>
10. [Philology.ru](http://Philology.ru)

## **7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

В учебном процессе используются методические рекомендации по изучению дисциплины, утвержденные УМС МГОУ:

1. Методические рекомендации по организации самостоятельной работы обучающихся, авторы Бондаренко И.В., Фильчакова Е.М.
2. Методические рекомендации по подготовке научного доклада, авторы Горбачева О.А., Харитоновна Е.Ю.
3. Методические рекомендации к написанию курсовой работы, авторы Беляева И.Ф., Савченко Е.П., Харитоновна Е.Ю.

## **8. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ**

### **Лицензионное программное обеспечение:**

Microsoft Windows  
Microsoft Office  
Kaspersky Endpoint Security

### **Информационные справочные системы:**

Система ГАРАНТ  
Система «КонсультантПлюс»

### **Профессиональные базы данных**

[fgosvo.ru](http://fgosvo.ru)  
[pravo.gov.ru](http://pravo.gov.ru)  
[www.edu.ru](http://www.edu.ru)

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

Материально-техническое обеспечение дисциплины включает в себя:

- учебные аудитории для проведения занятий лекционного и семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные учебной мебелью, доской, демонстрационным оборудованием;

- помещения для самостоятельной работы, укомплектованные учебной мебелью, персональными компьютерами с подключением к сети Интернет и обеспечением доступа к электронным библиотекам и в электронную информационно-образовательную среду МГОУ;
- помещения для хранения и профилактического обслуживания учебного оборудования, укомплектованные мебелью (шкафы/стеллажи), наборами демонстрационного оборудования и учебно-наглядными пособиями.