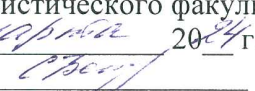


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталия Александровна
Должность: Ректор
Дата подписания: 27.06.2025 11:11:51
Уникальный программный ключ:
6b5279da4e034bffa79172803da5b7b559fc69e2

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ)

Лингвистический факультет
Кафедра теории языка, англистики и прикладной лингвистики

Согласовано
деканом лингвистического факультета
« 14 » марта 2024 г.


/Вековищева С.Н./

Рабочая программа дисциплины

Компьютерная лингвистика

Направление подготовки
45.03.02 Лингвистика

Профиль:
Цифровая лингвистика (английский язык + китайский или корейский языки)

Квалификация
Бакалавр

Форма обучения
Очная

Согласовано учебно-методической комиссией
лингвистического факультета
Протокол «14» марта 2024 г. № 5
Председатель УМКом 

/Горбачева О.А./

Рекомендовано кафедрой теории языка,
англистики и прикладной лингвистики
Протокол от «26» февраля 2024 г. № 8
Зав. кафедрой 

/Холстинина Т.В./

Мытищи
2024

Автор-составитель:

Максименко Ольга Ивановна, доктор филологических наук, профессор

Рабочая программа дисциплины «Компьютерная лингвистика» составлена в соответствии с требованиями Федерального Государственного образовательного стандарта высшего образования по направлению подготовки 45.03.02 Лингвистика утвержденного приказом МИНОБРНАУКИ РОССИИ от 12.08.2020, № 969

Дисциплина относится к части, формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)» и является обязательной для изучения.

Реализуется в формате электронного обучения с применением дистанционных образовательных технологий.

Год начала подготовки (по учебному плану) 2024

СОДЕРЖАНИЕ

1.ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ.....	Ошибка! Закладка не определена.
2.МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	Ошибка! За
3.ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	6
4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ	Ошибка! Закладка не определена.
5.ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ.....	9
6.УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	Ошибка! За
7.МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ	Ошибка! Закладка не оп
8.ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ	Ошибка! Закладка не определена.
9.МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	Ошибка! Закладка не

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

1.1. Цель и задачи дисциплины

Образовательная цель программы «Компьютерная лингвистика» 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.

Практическая цель – способствовать систематическому формированию соответствующих компетенций будущих бакалавров

Задачи дисциплины:

- формирование у студента представления об объекте, предмете и задачах основ компьютерной лингвистики, основных методах исследований дисциплины;
- формирование у студентов умения работать с научной литературой, аналитически осмысливать и обобщать теоретические положения;
- выработка умений по анализу структур конкретного языка на разных языковых уровнях, что может способствовать проведению общего анализа и синтеза для решения как общетеоретических, так и прикладных задач автоматической обработки естественного языка.

1.2. Планируемые результаты обучения

В результате освоения данной дисциплины у обучающихся формируются следующие компетенции:

СПК-1. Способен применять систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях о закономерностях функционирования языков мира

СПК-2. Способен использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач

СПК-3. Владеет основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических данных

СПК-4 . Способен осуществлять межъязыковое и межкультурное взаимодействие в устной и письменной формах как в общей, так и профессиональной сферах общения

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к части, формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)» и является обязательной для изучения.

Учебная дисциплина в методическом плане основывается на теоретических знаниях, полученных студентами в курсах дисциплин, таких как прикладная лингвистика, фонетика, понятийный аппарат математики, практический курс первого иностранного языка(английский язык).

Полученные в процессе обучения знания могут быть использованы при изучении таких дисциплин как «Морфология», «Синтаксис», «Семантика», «Лексикология», , «Дискурсивный анализ», «Лингвистическая экспертиза».

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:	82,7(48) ¹
Лекции	14(8) ²
Практические занятия	68(40) ³
Из них, в форме практической подготовки	68
Контактные часы на промежуточную аттестацию	0,7
Курсовая работа	0.3
Зачет с оценкой	0.4
Самостоятельная работа	100
Контроль	33,3

Форма промежуточной аттестации - 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	Количество часов		
	Лекции	Практические занятия	
		Общее количество	Из них, в форме практической подготовки
Тема 1. Компьютерная лингвистика: цели, задачи и методы. Современные приложения компьютерной лингвистики. Определение компьютерной лингвистики, связь с другими дисциплинами. История развития компьютерной лингвистики. Библиотеки	2	4	4

¹ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

² Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

³ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

инструментов для обработки текста.			
Тема 2. Первые этапы обработки текста. Необходимость обработки «сырого» текста. Токенизация. Другие этапы предобработки текста.	2	6(6) ⁵	6
Тема 3. Регулярные выражения. Понятие регулярных выражений. Метасимволы регулярных выражений. Применение регулярных выражений для обработки языка.	2(2) ⁴	6(6) ⁵	6
Тема 4. Компьютерная морфология. Основные понятия. Стемминг и лемматизация: алгоритмы и различия.		8(8) ⁵	8
Тема 5. Конечно-автоматная морфология. Конечный автомат как способ организации словаря для морфологического анализа и синтаксиса.	2(2) ⁵	2	2
Тема 6. Скрытые Марковские модели. Основная структура. Применение для обработки морфологии.		2	2
Тема 7. Задание правил для анализа морфологии. Типы правил для обработки фонологических особенностей. Создание набора правил для обработки морфологии. Теория оптимальности.		8(8) ⁵	8
Тема 8. Особенности создания лингвистических процессоров и обработки семантики и прагматики. Пирамида обработки естественного языка.	2	8	8
Тема 9. Особенности обработки синтаксиса. Порождающие грамматики. Исчисление Ламбека. SKY-парсинг.	2(2) ⁶	8	8
Тема 10. Представление текста и слова в виде вектора значений. Задание документа в виде вектора: способы, применение векторной модели. Дистрибутивная гипотеза. Word2vec.	2(2) ⁷	4	4
Тема 11. Компьютерные пакеты обработки текста.		12(12) ⁵	12
Итого	14(8)⁸	68(40)⁹	68

⁴ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

⁵ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

⁶ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

⁷ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

⁸ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

⁹ Реализуется в формате электронного обучения с применением дистанционных образовательных технологий

ПРАКТИЧЕСКАЯ ПОДГОТОВКА

Тема	Задание на практическую подготовку	количество часов
Тема2.Первые этапы обработки текста.	Провести процедуру токенизации и лемматизации заданного текста	6
Тема3. Регулярные выражения.	Проанализировать применение регулярных выражений для обработки языка	6
Тема4. Компьютерная морфология.	Стемминг и лемматизация: алгоритмы и различия. Осуществить данные процедура на заданном тексте	8
Тема 7. Задание правил для анализа морфологии.	Теория оптимальности. Анализ, подготовка аналитической справки	8
Тема 11. Компьютерные пакеты обработки текста.	Проанализировать и протестировать любую систему обработки текста https://soware.ru/categories/text-processing-systems	12
		40

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Темы для самостоятельного изучения	Изучаемые вопросы	Количество часов	Формы самостоятельной работы	Методическое обеспечение	Формы отчетности
Основы обработки текста	Предобработка, регулярные выражения	10	Чтение теории по обсуждаемым вопросам	Основная и дополнительная литература, интернет-ресурсы	Практическое задание
Библиотеки инструментов для анализа данных.	1.Natural Language ToolKit 2.Anaconda 3.Доступные парсеры и постеггеры: обзор и анализ системы.	16	Чтение теории по обсуждаемым вопросам	Основная и дополнительная литература, интернет-ресурсы	Практическое задание
Лемматизация	Алгоритмы	16	Чтение	Основная и	Практическое

и стемминг	лемматизации и стемминга Написание алгоритма стемминга для русского языка		теории по обсуждаемым вопросам	дополнительная литература, интернет-ресурсы	е задание,
Лексические преобразователи	Конечные автоматы Написание автомата для фрагмента морфологии русского языка	10	Чтение теории по обсуждаемым вопросам	Основная и дополнительная литература, интернет-ресурсы	Практическое задание
Формальные языки и грамматики	Формальное представление грамматики русского языка	16	Чтение теории по обсуждаемым вопросам	Основная и дополнительная литература, интернет-ресурсы	Практическое задание
Векторная модель	Представление документа в виде вектора Представление значения слова в виде вектора	16	Чтение теории по обсуждаемым вопросам	Основная и дополнительная литература, интернет-ресурсы	Практическое задание
Формальная семантика	Ассерция и презумпция Построение семантического дерева	16	Чтение теории по обсуждаемым вопросам	Основная и дополнительная литература, интернет-ресурсы	Практическое задание
Итого		100			

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

Код и наименование компетенции	Этапы формирования
СПК-1. 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
СПК-2. Способен использовать понятийный аппарат теоретической и прикладной	1. Работа на учебных занятиях 2. Самостоятельная работа

лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач	
СПК-3. 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
СПК-4. Способен осуществлять межъязыковое и межкультурное взаимодействие в устной и письменной формах как в общей, так и профессиональной сферах общения	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	Знать: основную лингвистическую информацию в рамках ПЛ Уметь: применять систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях о закономерностях функционирования языков мира	Практическое задание, тестирование доклада	Шкала оценивания практического задания Шкала оценивания тестирования Шкала оценивания доклада

	Продв инутый	1.Работа на учебных занятиях 2.Самостоятельна я работа	Знать: основную лингвистическую информацию в рамках ПЛ Уметь: применять систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях о закономерностях функционирования языков мира Владеть: систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях	Практическо е задание, тестировани е доклад, групповой/и ндивидуальн ый проект, практическа я подготовка	Шкала оценивания практическог о задания Шкала оценивания тестирования Шкала оценивания доклада Шкала оценивания группового/и ндивидуальн ого проекта Шкала оценивания практическо й подготовки
СПК-2	Порого вый	1.Работа на учебных занятиях 2.Самостоятельна я работа	Знать: основные понятия и методы изучаемых дисциплин Уметь: использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач	Практическо е задание, тестировани е доклад	Шкала оценивания практическог о задания Шкала оценивания тестирования Шкала оценивания доклада
	Продв инутый	1.Работа на учебных занятиях 2.Самостоятельна я работа	Знать: основные понятия и методы изучаемых дисциплин Уметь: использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач	Практическо е задание, тестировани е доклад, групповой/и ндивидуальн ый проект, практическа я подготовка	Шкала оценивания практическог о задания Шкала оценивания тестирования Шкала оценивания доклада Шкала оценивания группового/и ндивидуальн

			Владеть: навыками использования понятийного аппарата теоретической и прикладной лингвистики, для решения профессиональных задач		ого проекта Шкала оценивания практическо й подготовки
СПК-3	Порого вый	1.Работа на учебных занятиях 2.Самостоятельна я работа	Знать: основные математико-статистическими методами обработки лингвистической информации Уметь: работать с традиционными носителями информации, распределенными базами данных и знаний и другими электронными ресурсами	Практическо е задание, тестировани е доклад	Шкала оценивания практическог о задания Шкала оценивания тестирования Шкала оценивания доклада
	Продв инутый	1.Работан на учебных занятиях 2.Самостоятельна я работа	Знать: основные математико-статистическими методами обработки лингвистической информации Уметь: работать с традиционными носителями информации, распределенными базами данных и знаний и другими электронными ресурсами Владеть: основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических данных	Практическо е задание, тестировани е доклад, групповой/и ндивидуальн ый проект, практическа я подготовка	Шкала оценивания практическог о задания Шкала оценивания тестирования Шкала оценивания доклада Шкала оценивания группового/и ндивидуальн ого проекта Шкала оценивания практическо й подготовки
СПК-4	Порого	1.Работа на	Знать: основные	Практическо	Шкала

	вый	учебных занятиях 2.Самостоятельна я работа	характеристики межъязыкового и межкультурного взаимодействия в устной и письменной формах как в общей, так и профессиональной сферах общения Уметь: работать с глобальными компьютерными сетями.	е задание, тестировани е доклад	оценивания практическог о задания Шкала оценивания тестирования Шкала оценивания доклада
	Продв инутый	1.Работан на учебных занятиях 2.Самостоятельна я работа	Знать: основные характеристики межъязыкового и межкультурного взаимодействия в устной и письменной формах как в общей, так и профессиональной сферах общения Уметь: работать с глобальными компьютерными сетями. Владеть: способностью осуществлять межъязыковое и межкультурное взаимодействие в устной и письменной формах как в общей, так и профессиональной сферах общения	Практическо е задание, тестировани е доклад, групповой/и ндивидуальн ый проект, практическа я подготовка	Шкала оценивания практическог о задания Шкала оценивания тестирования Шкала оценивания доклада Шкала оценивания группового/и ндивидуальн ого проекта Шкала оценивания практическо й подготовки

Шкала оценивания тестирования

Критерии оценивания	Баллы
Выполнено правильно как минимум 80% заданий	26 баллов
Выполнено правильно как минимум 60% заданий	23 балла
Выполнено правильно как минимум 40% заданий	16 баллов
Выполнено правильно менее 40% заданий	12 баллов

2. Шкала оценивания выполнения практического задания

Критерии оценивания	Баллы
Выполнено правильно как минимум 80% предложенного задания	26 баллов
Выполнено правильно как минимум 60% предложенного задания	22 балла
Выполнено правильно как минимум 40% предложенного задания	18 баллов
Выполнено правильно менее 40% предложенного задания	12 баллов

Шкала оценивания доклада

Критерии оценивания	Баллы
Представленный доклад свидетельствует о проведенном самостоятельном исследовании с привлечением различных источников информации; во время выступления продемонстрированы коммуникативные и рефлексивные умения.	28 баллов
Представленный доклад свидетельствует о проведенном исследовании с привлечением различных источников информации, однако исследование выполнено не самостоятельно, допущены ошибки в изложении материала	22 балла
Представленный доклад свидетельствует о проведенном исследовании, однако слабо отражает знакомство с предметом/не соответствует теме исследования, исследование выполнено не самостоятельно, допущены ошибки в изложении материала	18 баллов
Представленный доклад не отражает знакомство с предметом/не соответствует теме исследования, исследование выполнено не самостоятельно, допущены ошибки в изложении материала	12 баллов

Шкала оценивания групповых/индивидуальных проектов

Критерии оценивания	Баллы
Выставляется, если представленный на защите продукт свидетельствует о проведенном самостоятельном исследовании с привлечением различных источников информации, отражает теоретическую и практическую направленность проекта; во время защиты проекта продемонстрированы коммуникативные и рефлексивные умения, а также навыки работы в команде.	16 баллов
Выставляется, если представленный на защите продукт свидетельствует о проведенном исследовании с привлечением различных источников информации, отражает теоретическую и практическую направленность	15 баллов

проекта.	
Выставляется, если представленный на защиту продукт свидетельствует о проведенном исследовании, отражает теоретическую направленность проекта.	9 баллов
Выставляется, если представленный на защиту продукт свидетельствует о проведенном исследовании с привлечением одного источника информации; слабо отражает теоретическую и практическую направленность проекта.	7 баллов

Шкала оценивания заданий практической подготовки

Критерии оценивания	Баллы
Выполнено правильно как минимум 80% предложенного задания	26 баллов
Выполнено правильно как минимум 60% предложенного задания	22 балла
Выполнено правильно как минимум 40% предложенного задания	18 баллов
Выполнено правильно менее 40% предложенного задания	12 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

Примерные задания по практической подготовке

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

I want the dog to eat my homework.

Посчитайте расстояние Левенштейна для пары слов:

Распределение – уравнение

```
(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))))))
    (. .)))
```

-
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 эмотиконов
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

Примерный список 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

Примерные темы индивидуальных и групповых проектов

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.

Примерные вопросы к зачету с оценкой

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.

В рамках освоения дисциплины предусмотрены следующие формы текущего контроля: выполнение практического задания, тестирование, доклад групповой/индивидуальный проект, практическая подготовка.
Общее количество баллов по дисциплине – 100 баллов.

Максимальное количество баллов, которое может набрать обучающийся в течение семестра за текущий контроль, равняется 70 баллам.

Промежуточная аттестация проводится устно по вопросам в форме зачета с оценкой. На зачете с оценкой студент должен дать развернутые ответы на теоретические вопросы. Максимальное число баллов, которые выставляются студенту по итогам по итогам зачета с оценкой – 30 баллов.

Требования к промежуточной аттестации

При оценке знаний на **зачете с оценкой** 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 зачета с оценкой

Критерии оценивания	Баллы
Выставляется за ответ, который демонстрирует прекрасное знание предмета, умение соединять знания из различных разделов курса, легко и безошибочно	30-21 балл

иллюстрировать теоретические положения примерами, как взятыми из учебника, так и своими собственными; владение терминологией из различных разделов курса.	
Выставляется за ответ, который демонстрирует хорошее знание и понимание изученного материала, подкреплён примерами, взятыми из лекций или учебника; допускаются единичные ошибки, которые экзаменуемый исправляет самостоятельно после замечаний преподавателя.	20-16 баллов
Выставляется за ответ, который обнаруживает самое общее понимание теории, однако, плохо подкрепляемое практическими примерами. При таком ответе студент проявляет неуверенность, не всегда даёт исчерпывающие аргументированные ответы на заданные вопросы	15-11 баллов
Выставляется за ответ, который обнаруживает непонимание сути вопроса, являясь механическим повторением курса лекций или учебника; незнание терминологии, искажение смысла понятий; неумение соотнести теорию с практикой.	10-0 баллов

Итоговая шкала выставления оценки по дисциплине

Итоговая оценка складывается из оценки за выполнения всех предусмотренных в программе дисциплины форм отчетности в рамках текущего контроля, а также оценки на промежуточной аттестации.

Баллы, полученные в течение освоения дисциплины	Оценка по дисциплине
81-100	отлично
61-80	хорошо
41-60	удовлетворительно
0-40	неудовлетворительно

Критерии оценки курсовых работ

1. Актуальность проведенного исследования.
2. Соответствие содержания работы заявленной теме исследования.
3. Степень самостоятельности.
4. Глубина проработки материала.
5. Правильность и полнота разработки поставленных вопросов.
6. Логичность и последовательность изложения.
7. Результативность выполненного практического исследования.
8. Значимость выводов для последующей практической деятельности.
9. Качество использованных методических источников.
10. Уровень грамотности.
11. Соответствие оформления предписанным стандартам.
12. Соблюдение сроков выполнения работы.
13. Проверка системой Антиплагиат.

Шкала оценивания курсовых работ

– оценка «отлично» **81-100 баллов** выставляется студенту, если курсовая работа представляет собой самостоятельный анализ нескольких источников по конкретной теме,

работа свидетельствует о навыках анализа и обобщения изученного материала, а также собственном мнении студента об излагаемой проблеме и умении правильно оформить и защитить работу;

- оценка **«хорошо» 61-80 баллов** выставляется студенту, если курсовая работа представляет собой самостоятельный анализ нескольких источников по конкретной теме, однако при изложении изученного материала отмечаются отдельные нарушения логического обобщения, а также наблюдаются неточности в оформлении работы и при ее защите;

- оценка **«удовлетворительно» 41-60 баллов** выставляется студенту, если в курсовой работе представлен анализ недостаточного количества источников по конкретной теме, при изложении изученного материала отмечаются значительные нарушения логического обобщения, а также имеются ошибки в оформлении работы и при ее защите;

– оценка **«неудовлетворительно» 0-40 баллов** выставляется студенту в том случае, если автор не обнаруживает навыки обобщения и логического изложения исследуемого материала или курсовая работа не является самостоятельно проведенным анализом.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1 Основная литература:

1. Гуслякова, А. В. Информационные технологии и лингвистика XXI века : учебное пособие / А. В. Гуслякова. – 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. Козубинский ; 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»

Электронно-библиотечная система Лань – <https://e.lanbook.com/>
Электронно-библиотечная система IR books – <https://www.iprbookshop.ru/>
Электронно-библиотечная система Консультант студента – <https://www.studentlibrary.ru/>
Электронно-библиотечная система Университетская библиотека онлайн – <https://biblioclub.ru/>
Электронно-библиотечная система Юрайт – <https://urait.ru/>
EastViewUniversalDatabases – <http://www.ebiblioteka.ru/>. –
Библиотека – http://www.mgou.ru/index.php?option=com_content&task=view&id=204&Itemid=199.
Электронно-библиотечная система Znanium.com – <http://www.znaniy.com/>.
http://terralinguistica.ru/harbour/Typology_materials/Typology_literature.html
<http://www.philology.ru>
<http://www.lomonosov-fund.ru>
http://www.gramota.ru/spravka/docs/16_7
www.superlinguist.ru
http://genhis.philol.msu.ru/printer_206.html

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

1. Методические рекомендации по организации самостоятельной работы студентов
2. Методические рекомендации по написанию курсовой работы

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 «Консультант Плюс»

Профессиональные базы данных

fgosvo.ru - [Портал Федеральных государственных образовательных стандартов высшего образования](#)

pravo.gov.ru - [Официальный интернет-портал правовой информации](#)

www.edu.ru - Федеральный портал Российское образование

Свободно распространяемое программное обеспечение, в том числе отечественного производства

ОМС Плеер (для воспроизведения Электронных Учебных Модулей)

7-zip

Google Chrome

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.