

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет романо-германской филологии

УТВЕРЖДАЮ
Проректор по учебной работе,
качеству образования – первый
проректор

Харузов Т.А.
подпись
« 28 » _____ 20 21 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.В.ДВ.03.02 Компьютерная лингвистика

Направление подготовки 45.03.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) Связь, информационные и коммуникационные технологии в сфере управления информационными ресурсами

Форма обучения очная

Квалификация бакалавр

Краснодар 2021

Рабочая программа дисциплины Компьютерная лингвистика составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 45.03.03 Фундаментальная и прикладная лингвистика

Программу составил(и):

Колчевская В.А., преподаватель кафедры прикладной лингвистики и новых информационных технологий


подпись

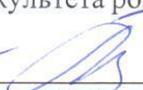
Рабочая программа дисциплины Компьютерная лингвистика утверждена на заседании кафедры прикладной лингвистики и новых информационных технологий протокол № 7 «14» мая 2021г.

Заведующий кафедрой (разработчик) Бодоньи М.А.


подпись

Утверждена на заседании учебно-методической комиссии факультета романо-германской филологии протокол № 5 «18» мая 2021г.

Председатель УМК факультета Бодоньи М.А.


подпись

Рецензенты:

Кулинцева Н.А., канд. филол. наук, доцент кафедры западноевропейских языков и культур ФГБОУ ВО «Пятигорский государственный университет»

Зиньковская А.В., д-р филол. наук, заведующий кафедрой английской филологии ФГБОУ ВО «Кубанский государственный университет»

1. Цель и задачи изучения дисциплины (модуля)

1.1 Цель освоения дисциплины

Цель данной дисциплины – дать студентам общее представление о задачах компьютерной лингвистики, а также заложить базу для дальнейшего освоения курсов, относящихся к теории и практике языковых технологий

1.2 Задачи дисциплины

В процессе освоения дисциплины реализуются следующие задачи:

- способствовать теоретическому осмыслению и систематизации основных понятий и категории современной лингвистики, а также основных направлений и задач компьютерной лингвистики;
- изучить основные типы систем, использующих модули лингвистического анализа;
- овладеть основными методами, способами и средствами получения, хранения, переработки информации;
- расширять знания студентов об типах, характеристиках и особенностях основных доступных в Интернете лингвистических ресурсов;
- сформировать умения использовать основные принципы и методы компьютерного моделирования лингвистических задач.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Компьютерная лингвистика» относится к части, формируемой участниками образовательных отношений Блока 1 "Дисциплины (модули)" учебного плана. Предшествующими дисциплинами для освоения дисциплины «Компьютерная лингвистика» являются Б1.В.09 Конструирование веб-ресурсов Б1.В.ДВ.03.01 Корпусная лингвистика; последующие дисциплины: Б1.В.ДВ.04.02 Компьютерная лексикография и ФТД.01 Интернет-ресурсы в обучении иностранному языку.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующей компетенции

Код и наименование индикатора	Результаты обучения по дисциплине
ПК-3 Способен пользоваться лингвистически ориентированными программными продуктами	
ПК-3.1 Имеет представления о лингвистически ориентированных программных продуктах	Знает понятия и представления лингвистически ориентированных программных продуктов
	Умеет применить знания для создания новых лингвистически ориентированных программных продуктов
	Владеет способностью к разработке лингвистически ориентированных программных продуктов

Результаты обучения по дисциплине достигаются в рамках осуществления всех видов контактной и самостоятельной работы обучающихся в соответствии с утвержденным учебным планом.

Индикаторы достижения компетенций считаются сформированными при достижении соответствующих им результатов обучения.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 3 зачетных единиц (108 часов), их распределение по видам работ представлено в таблице:

Виды работ	Всего часов	Форма обучения
		очная
		6 семестр (часы)
Контактная работа, в том числе:	28,3	28,3
Аудиторные занятия (всего):	26	26
занятия лекционного типа	12	12
лабораторные занятия	14	14
практические занятия	-	-
семинарские занятия	-	-
Иная контактная работа:	2,3	2,3
Контроль самостоятельной работы (КСР)	2	2
Промежуточная аттестация (ИКР)	0,3	0,3
Самостоятельная работа, в том числе:	44	44
Самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам и т.д.)	30	30
Подготовка к текущему контролю	14	14
Контроль:		
Подготовка к экзамену	35,7	35,7
Общая трудоемкость	час.	108
	в том числе контактная работа	28,3
	зач. ед	3

2.2 Содержание дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 5 семестре (3 курсе) (очная форма обучения)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	СРС
1	Введение. Предмет компьютерной лингвистики. Области исследования. Подходы. История.	14	2		2	10
2	Основные направления автоматической обработки текстов. Теоретические и прикладные направления автоматической обработки текстов. Машинная морфология.	14	2		2	10
3	Основные типы компьютерных лингвистических ресурсов. Корпуса текстов. Лингвистическая дешифровка.	14	2		2	10
4	Компьютерные словари, тезаурусы, онтологии. Моделирование в компьютерной лингвистике.	9	2		2	5
5	Язык регулярных выражений. Понятие регулярного выражения. Синтаксис регулярных выражений. Использование языка регулярных выражений для разбиения текста на слова. Использование языка регулярных выражений для поиска в тексте	9	2		2	5
6	Автоматический морфологический анализ и применение конечных преобразователей морфологическом анализе. Основные проблемы автоматического семантического анализа.	10	2		4	4
	<i>ИТОГО по разделам дисциплины</i>	70	12		14	44
	Контроль самостоятельной работы (КСР)	2				2
	Промежуточная аттестация (ИКР)	0.3				0.3
	Подготовка к текущему контролю	35,7				35,7
	Общая трудоемкость по дисциплине	108				108

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1 Занятия лекционного типа.

№	Наименование раздела (темы)	Тематика занятий/работ	Форма текущего контроля
---	-----------------------------	------------------------	-------------------------

1	Введение. Предмет компьютерной лингвистики. Области исследования. Подходы. История.	О термине «информатика». Область возникновения лингвистических проблем информатики. Совершенствование массовой и индивидуальной коммуникации. Информатика и компьютерная лингвистика.	—
2	Основные направления автоматической обработки текстов. Теоретические и прикладные направления автоматической обработки текстов. Машинная морфология.	Ввод текста в компьютер. Человеко-компьютерное взаимодействие. Компьютерная лингвистика. Обработка лингвистической информации на уровне словоформ, слов, словосочетаний, предложений, текста.	—
3	Основные типы компьютерных лингвистических ресурсов. Корпуса текстов. Лингвистическая дешифровка.	Автоматический морфологический анализ. Виды автоматического морфологического анализа. Современное состояние морфологического анализа.	—
4	Компьютерные словари, тезаурусы, онтологии. Моделирование в компьютерной лингвистике.	Лингвистическая дешифровка как прикладная дисциплина. Статистические методы. Графематический уровень. Дериватология.	—
5	Язык регулярных выражений. Понятие регулярного выражения. Синтаксис регулярных выражений. Использование языка регулярных выражений для разбиения текста на слова. Использование языка регулярных выражений для поиска в тексте	Традиционная и машинная лексикография. Отличия машинного словаря от обычного. Вычислительная лексикография. Словарно-центрический подход. Лемматизация. Машиночитаемые словари. Лемматизация. Составление машинных словарей.	—
6	Автоматический морфологический анализ и применение конечных преобразователей морфологическом анализе. Основные проблемы автоматического семантического анализа.	Моделирование языковых сущностей и человеческого мышления. Связь языка с мышлением. Элементы системы искусственного интеллекта. Как мыслит человек. Искусственный интеллект. Модель механизма мышления.	—

2.3.2 Занятия семинарского типа (лабораторные работы)

№	Наименование раздела (темы)	Тематика занятий/работ	Форма текущего контроля
---	-----------------------------	------------------------	-------------------------

1.	Введение. Предмет компьютерной лингвистики. Области исследования. Подходы. История.	Что такое компьютерная лингвистика? Основные направления компьютерной лингвистики.	Контрольная работа №1
2.	Основные направления автоматической обработки текстов. Теоретические и прикладные направления автоматической обработки текстов. Машинная морфология.	Словари. Корпуса текстов.	Реферат
3.	Основные типы компьютерных лингвистических ресурсов. Корпуса текстов. Лингвистическая дешифровка.	Компьютерный анализ текста. Задачи лингвистических информационных технологий.	Контрольная работа № 1
4.	Компьютерные словари, тезаурусы, онтологии. Моделирование в компьютерной лингвистике.	Морфологический уровень. Синтаксический уровень. Анафора и кореферентность.	Тест
5.	Язык регулярных выражений. Понятие регулярного выражения. Синтаксис регулярных выражений. Использование языка регулярных выражений для разбиения текста на слова. Использование языка регулярных выражений для поиска в тексте	Закон Ципфа. Модель TF*IDF. Классификация документов. Классификация с обучением. Оценка результатов классификации.	Проект
6.	Автоматический морфологический анализ и применение конечных преобразователей морфологическом анализе. Основные проблемы автоматического семантического анализа.	Кластеризация. Контент-анализ.	Контрольная работа № 2

2.3.3 Практические занятия.

Практические занятия – не предусмотрены

2.3.4 Примерная тематика курсовых работ (проектов).

Курсовые работы – не предусмотрены

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю).

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
---	---------	---

1	Самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам и т.д.)	Методические указания по организации самостоятельной работы по дисциплине «Компьютерная лингвистика», утвержденные кафедрой прикладной лингвистики и новых информационных технологий, протокол № от «» мая 2021 г.
---	---	--

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа;

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии.

В рамках лабораторных занятий студентам предлагается подготовить рефераты и эссе по интерпретации сочетаемости слов в русских и зарубежных словарях, что предполагает развитие навыка самостоятельной работы с научной литературой. В ходе обучения активно используются задания не только на основе отдельных слов, но и на основе целых словарей, что способствует усвоению не только языковых теоретических особенностей системы, но их речевых воплощений. Студентам предлагаются проблемные вопросы, предполагающие различные решения с позиций разных концепций, что развивает способность учитывать многогранность языковых и речевых явлений и уметь аргументировано отстаивать свою позицию в научном споре, уметь принимать во внимание контраргументы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4.1 Фонд оценочных средств для проведения текущего контроля.

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Компьютерная лингвистика».

Оценочные средства включают контрольные материалы для проведения **текущего контроля** в форме тестовых заданий, реферата, проекта, контрольной работы) и **промежуточной аттестации** в форме вопросов к зачету.

Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Код и наименование индикатора	Результаты обучения	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация
1	ПК-3.1 Имеет представления о лингвистически	Знает понятия и представления лингвистически	Контрольная Работа № 1	Вопросы на экзамене 1-20

	ориентированных программных продуктах	ориентированных программных продуктов		
2	ПК-3.1 Имеет представления о лингвистически ориентированных программных продуктах	Умеет применить знания для создания новых лингвистически ориентированных программных продуктов	Реферат, тест, проект	Вопросы на экзамене 21-40
3	ПК-3.1 Имеет представления о лингвистически ориентированных программных продуктах	Владеет способностью к разработке лингвистически ориентированных программных продуктов	Контрольная Работа № 2	Вопрос на зачете 41-50

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

1) Тест

В каждом задании — 1 правильный ответ, за каждый правильный ответ дается 1 балл

1. Какое из высказываний является определением прикладной лингвистики?

- a) область языкознания, направленная на объективное установление состояния отдельного языка, его истории и закономерностей;
- b) область языкознания, связанная с использованием компьютерных инструментов — программ, технологий организации и обработки данных — для моделирования функционирования языка в тех или иных условиях;
- c) область языкознания, связанная с разработкой методов решения практических задач использования языка;
- d) область языкознания, связанная с применением компьютерных моделей языка в лингвистике и в смежных с ней дисциплинах.

2. К направлениям компьютерной лингвистики не относится

- a) компьютерная лексикография;
- b) компьютерно-опосредованная коммуникация;
- c) системы обработки естественного языка;
- d) машинный перевод.

3. Информатика — это

- a) наука об управлении, связи и переработке информации;
- b) наука о накоплении, обработке и передаче информации с помощью ЭВМ;
- c) наука о накоплении, обработке и передаче информации о строении языка с помощью ЭВМ;
- d) наука об использовании компьютерных инструментов для моделирования функционирования языка в тех или иных условиях.

4. Разное количество информации в одном и том же сообщении для разных людей зависит не от...

- a) накопленных ими знаний;
- b) уровня понимания сообщения;
- c) их интереса к сообщению;

- d) их уровня владения компьютерной техникой.
5. Следствие третьей информационной революции состоит в том, что...
- a) информация становится общедоступной;
 - b) информацию можно автоматически обрабатывать и передавать с большой скоростью;
 - c) информацию можно легко найти с помощью инструментов поиска и совместно производить;
 - d) информация может накапливаться.
6. Для современного человека преобладающей является...
- a) звуковая информация;
 - b) визуальная (символьная) информация;
 - c) вкусовая и тактильная информация;
 - d) визуальная (образная) информация.
7. Адекватность информации — это ...
- a) степень соответствия информации объективной реальности окружающего мира;
 - b) степень соответствия информации, полученной потребителем, тому, что автор вложил в ее содержание;
 - c) достаточность информации для принятия решения;
 - d) степень соответствия информации текущему моменту времени.
8. Машинный синтаксис — это ...
- a) правила строения имен;
 - b) правила построения слов в более сложные структуры;
 - c) соотношение слова и его значения;
 - d) правила перевода письменного символа в устный.
9. Естественный язык — это ...
- a) знаковая система, используемая человеком с момента рождения;
 - b) знаковая система, используемая человеком в непринужденной обстановке;
 - c) знаковая система, созданная для естественных наук;
 - d) знаковая система, стихийно возникшая и закрепившаяся в обществе.
10. Волапюк — это...
- a) специализированный язык науки;
 - b) родной язык одного из малочисленных племен;
 - c) неспециализированный искусственный язык;
 - d) система символического кодирования.
11. Какие из следующих приложений не являются текстовыми редакторами?
- a) MS Excel;
 - b) Corel WordPerfect;
 - c) MS Works;
 - d) Adobe InCopy.
12. Microsoft Word не включает...
- a) функции настольных издательских систем;
 - b) функцию удалённого доступа;
 - c) функцию редактирования графических объектов;
 - d) шаблоны типовых таблиц.
13. К устройствам ввода данных не относится
- a) сканер;
 - b) принтер;
 - c) клавиатура;
 - d) цифровой фотоаппарат.
14. OCR — это ...
- a) система автоматического распознавания символов;
 - b) система переводческой памяти;
 - c) система машинного перевода;

d) функция текстового процессора.

15. Реферат — это...

- a) связный текст, который кратко выражает тему, предмет, цель, методы и результаты исследования;
- b) процесс составления содержания документа (книги, статьи, патента на изобретение и др.);
- c) краткое изложение содержания документа, дающее общее представление о его теме;
- d) краткий текст, выполняющий сигнальную функцию (информирует о том, что есть публикация на определенную тему).

16. Слово, относящееся к основному содержанию текста и повторяющееся в нем несколько раз, в автоматическом реферировании называется ...

- a) лейтмотивом;
- b) термином;
- c) символом;
- d) ключевым словом.

17. Метод автоматического аннотирования, при котором важные слова выделяются в заголовке, подзаголовке, начале и конце текста, называется ...

- a) статистическим;
- b) логико-семантическим;
- c) позиционным;
- d) функциональным.

18. Совокупность специально отобранных текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска, называется ...

- a) базой данных;
- b) словарем;
- c) информационным массивом;
- d) корпусом.

19. Разметка бывает ...

- a) морфологической; синтаксической; семантической и просодической;
- b) полнотекстовой и фрагментной;
- c) синхронической и диахронической;
- d) звуковой, письменной, смешанной.

20. УНК — это...

- a) корпус естественного языка, представительный по отношению ко всему языку;
- b) универсальный национальный код;
- c) собрание текстов, которое существует в Интернете;
- d) собрание текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска.

21. Требования к корпусам

- a) полнота, адекватность, актуальность, компьютерная поддержка;
- b) устойчивость, тиражируемость, адаптируемость, оптимальность временных параметров, комфорт пользователя;
- c) репрезентативность, полнота, экономичность, структуризация, компьютерная поддержка;
- d) полнота, экономичность, достоверность, структуризация, компьютерная поддержка.

22. Корпусный менеджер ...

- a) обеспечивает сортировку результатов поиска, статистические подсчеты, составление списков слов на основе корпуса;
- b) это специальная программа поиска по корпусу;
- c) это человек, составляющий корпуса и управляющий ими;
- d) это специальная программа подготовки текстов к их включению в корпус.

23. ПОД — это ...

- a) вид информационно-поисковой системы;
 - b) специальная программа поиска по корпусу;
 - c) поисковый образ документа;
 - d) поисковая оценка данных.
24. Одна из основных проблем компьютерного анализа речи состоит в том, что ...
- a) невозможно создать искусственный интеллект;
 - b) компьютер не умеет работать со смыслом;
 - c) у компьютера нет дополнительных источников информации (ситуация, контекст, прошлый опыт в данной области и т.п.);
 - d) разработчики не желают делиться своими профессиональными секретами.
25. Электронный словарь — это ...
- a) введенный в компьютер бумажный словарь, снабженный средствами поиска и отображения информации;
 - b) организованное собрание слов с комментариями, в которых описываются особенности структуры и/или функционирования этих слов;
 - c) организованное собрание слов с описанием их значения, особенностей употребления, структурных свойств, сочетаемости, соотношения с лексическими системами других языков и т.д.;
 - d) словарь в специальном машинном формате, предназначенный для применения на ЭВМ пользователем или компьютерной программой.
26. К зонам словарной статьи не относится
- a) лексический вход (вокабула, лемма);
 - b) зона грамматической информации;
 - c) зона стилистических помет;
 - d) словник.
27. Что включает в себя понятие АСПОТ?
- a) словарь в специальном машинном формате, предназначенный для применения на ЭВМ пользователем;
 - b) компьютерные версии хорошо известных словарей (Вебстер, Коллинз, Ожегов...);
 - c) словарь в специальном машинном формате, предназначенный для применения на ЭВМ компьютерной программой;
 - d) словари, предназначенные для обычного пользователя.
28. Что не относится к понятию термина?
- a) слово (словосочетание) метаязыка науки, а также областей конкретной практической деятельности человека;
 - b) понятие задается через свойства, реализуемые в системе;
 - c) использование основывается не на интуиции, а на четких определениях;
 - d) сопоставляется, как правило, несколько значений.
29. Что не относится к процессу и понятию машинного перевода?
- a) междисциплинарность;
 - b) использование машинных средств;
 - c) принципиальное сходство этапов понимания и синтеза текста;
 - d) учет языковых и экстралингвистических знаний.
30. Типовая парадигма лексемы в автоматическом морфологическом анализе — это ...
- a) последовательность букв от начала словоформы, общая для всех словоформ;
 - b) элементы, описывающие формоизменение конкретной лексемы,
 - c) совокупность наборов машинных окончаний;
 - d) совпадение основ разных слов.

Критерии оценки теста.

Оценка по тесту выставляется пропорционально доле правильных ответов:
90-100% – оценка «отлично»

75-89% – оценка «хорошо»

60-74% – оценка «удовлетворительно»

менее 60% правильных ответов – оценка «неудовлетворительно»

2) Контрольная работа

1. *Определите статистические показатели приведенного ниже текста смешанного языкового типа.*

Проекты Sibola/Oleada реализуют обширные компьютерные системы лингвистического анализа текстов, представленных в Unicode. Компоненты системы включают средства работы с мультязыковыми текстами (MUTT), построения конкорданса (XConcord) для текстов на более чем 16 языках, статистического анализа, автоматического перевода, различные словари и тезаурусы. Некоторые версии этих компонентов доступны для бесплатной загрузки после процедуры формальной регистрации. Все компоненты реализованы в среде X11 Window System для SunOs и Solaris (источник: Проекты Sibola/Oleada <http://rvb.ru/soft/catalogue/c01.htm>)

Слов

Символов (без пробелов)

Символов (с пробелами)

Символов в латинской графике

Чисел

Средняя длина слов

2. *Какому языку соответствует средняя длина слов текста смешанного типа, приведенного в задании 1? Для выполнения задания вычислите среднюю длину слов русского языка из приведенного текста и среднюю длину слов в латинской графике.*

3. *Определите, каким языкам соответствуют следующие специфические буквы, буквосочетания и слова:*

а) 0 б В р со ёis,

б) th sch sc.

в) et, the, der, och, icin.

4. *Создайте диагностический словарь для определения языка на материале текстов на двух разных языках (на ваш выбор). Для этого заполните следующую таблицу:*

Критерий

Язык 1:

Язык 2:

Типичные артикли

Указательные местоимения

Местоимения 3-го лица

Отдельные формы

вспомогательных глаголов

Основные предлоги и союзы

Другие частотные слова

5. *Дополните таблицу встречаемости букв в распространенных европейских языках добавив в нее данные по русскому языку. Используйте для этого любой текст на русском языке объемом не менее 100 символов.*

6. *Прочитайте несколько фраз на эсперанто. Назовите морфологические диагностические показатели этого языка, учитывая, что существительные прилагательные на эсперанто всегда имеют одни и те же окончания.*

Рус. Эсперанто

зеленое дерево verda arbo

старый человек maljuna viro

хороший друг bela amiko

7. Найдите лишнее в приведенном ниже списке. Решите данную задачу с точки зрения компьютерной семантики и компьютерной грамматики.

Ландыш, левкой, лаватера, лютик, люпин, ромашка, липа.

8. Определите, к какому виду прикладных программ относятся перечисленные ниже программные продукты.

- 1) Текстовые редакторы
- 2) Графические редакторы
- 3) Электронные таблицы
- 4) Веб-редакторы
- 5) Веб-браузеры

Opera, MS Excel, MS FrontPage, Adobe Photoshop, Corel WordPerfect

Критерии оценки контрольной работы.

Оценка по тесту выставляется пропорционально доле правильных ответов:

90-100% – оценка «отлично»

75-89% – оценка «хорошо»

60-74% – оценка «удовлетворительно»

менее 60% правильных ответов – оценка «неудовлетворительно»

3) Реферат

1. Обзор сетевых ресурсов по корпусной лингвистике
2. Характеристика ресурсов по компьютерной лингвистике (www.dialog-21.ru, www.computer.org)
3. Специальные возможности программы MS Word для лингвистов (проверка правописания, рецензирование, автореферирование, использование шаблонов и т.д.)
4. Правильное использование заимствованных терминов и обозначений (правописание, склонение, спряжение, ударение) компьютерной лингвистики
5. Особенности электронных переводческих словарей Lingvo nMultitran и их отличия от онлайн-переводчиков (Google, Yandex и т.п.)
6. Сравнение программ переводческой памяти (TRADOS, Deja vu и т.п.)
7. Сравнение программ автоматического перевода (ПРОМТ, Сократ и т.п.)
8. Средства обеспечения и поддержки локализации (Multilizer, Passolo и т.п.)
9. Краудсорсинг или модель «Википедии» в переводе
10. Сравнение мультимедийных программ по обучению иностранным языкам (English DeLuxe, «РЕПЕТИТОР English) и т.п.)

Критерии оценки рефератов

Оценка 5 ставится, если выполнены все требования к написанию и защите реферата: обозначена проблема и обоснована её актуальность, сделан краткий анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция, сформулированы выводы, тема раскрыта полностью, выдержан объём (12-15 стр.), соблюдены требования к внешнему оформлению, даны правильные ответы на дополнительные вопросы.

Оценка 4 – основные требования к реферату и его защите выполнены, но при этом допущены недочёты. В частности, имеются неточности в изложении материала; отсутствует логическая последовательность в суждениях; не выдержан объём реферата (12-15 стр.); имеются упущения в оформлении; на дополнительные вопросы при защите даны неполные ответы.

Оценка 3 – имеются существенные отступления от требований к реферированию. В частности, тема освещена лишь частично; допущены фактические ошибки в содержании реферата или при ответе на дополнительные вопросы; во время защиты отсутствует вывод.

Оценка 2 – тема реферата не раскрыта, обнаруживается существенное непонимание проблемы либо реферат студентом не представлен.

4) Проект

Найдите в Интернете текст Alice's Adventures in Wonderland by Lewis Carroll (например, на сайте www.gutenberg.org/ebooks/11). Сохраните его на свой компьютер в формате MS Word. Выполните задания на простой поиск в этом документе и внесите результаты поиска в таблицу.

Задание

1. Сколько раз в тексте встречается слово child (в разных формах)?
2. Сколько раз в тексте встречается слово child именно в этой форме?
3. Приведите один из контекстов использования в тексте слова beautiful
4. В какой орфографии (британской или американской) представлен текст?

Поиск с подстановочными знаками

Выполните поиск с подстановочными знаками по тексту Alice's Adventures in Wonderland. Внесите результаты поиска в таблицу

Задание

Формула поиска

Ответ

1. Найдите в тексте первые пять слов, состоящих из пяти букв
2. Сколько в тексте шестибуквенных слов, начинающихся на букву s и заканчивающиеся на букву g?
3. Найдите в тексте первые пять трёхбуквенных слов, начинающиеся на гласную букву
4. Сколько в тексте слов, состоящих из двенадцати букв? По каким формальным признакам их можно сгруппировать? Приведите пример из каждой группы слов.

Ответ:

Группы:

5. Сколько в тексте слов с суффиксом -tion?

Приведите пример использования такого слова в контексте

Ответ:

Пример:

6. Есть ли в тексте слова, включающие четыре согласные буквы подряд?
7. Сколько раз в тексте встречаются пассивные конструкции единственного числа прошедшего времени?

Критерии оценки проекта

Каждый из предложенных показателей оценивается по критериям

выполнен – 2 балла

частично выполнен – 1 балл

не выполнен – 0 баллов.

Показатели оценки	Критерии оценивания
1	Структура (количество слайдов соответствует содержанию и продолжительности выступления, например: для 7-минутного выступления рекомендуется использовать не более 10 слайдов, включая титульный слайд и слайд с выводами)
2	Наглядность (иллюстрации хорошего качества, с четким изображением, текст легко читается, например: используются средства наглядности информации в виде таблиц, схем, графиков и т. д.)

3	Дизайн и настройка (оформление слайдов соответствует теме, не препятствует восприятию содержания, для всех слайдов презентации используется один и тот же шаблон оформления)
4	Содержание (презентация отражает основные этапы исследования – проблему, цель, гипотезу, ход выполнения работы, выводы, т.е. содержит полную, понятную информацию по теме доклада при наличии орфографической и пунктуационной грамотности)
5	Требования к выступлению (выступающий свободно владеет содержанием, ясно и грамотно излагает материал, выступающий свободно и корректно отвечает на вопросы и замечания аудитории, выступающий точно укладывается в рамки регламента).

Зачетно-экзаменационные материалы для промежуточной аттестации (экзамен)

- 1 Компьютерная лингвистика как направление научной деятельности, ее цели и задачи.
- 2 Связь компьютерной лингвистики с другими областями человеческого знания.
- 3 Искусственный интеллект как область знаний. Основные направления исследований.
- 4 Лингвистические базы знаний. Принципы формирования и использования.
- 5 Структура лингвистической базы знаний.
- 6 Соотношение прикладной лексикографии с общей лексикологией и другими смежными дисциплинами. Задачи прикладной лексикографии.
- 7 Словарь словоформ, его достоинства и недостатки, принципы разработки.
- 8 Словарь типа «машинная основа + машинная флексия», его достоинства и недостатки, принципы разработки
- 9 Концепция матричного гнездового (словообразовательного) словаря как один из способов снятия неполноты словарей.
- 10 Системное кодирование информации различного рода на уровне словаря. Основные требования к кодам.
- 11 Системное кодирование морфологической информации о именных частях речи в системах автоматической обработки текстов.
- 12 Системное кодирование морфологической информации о глаголе в системах автоматической обработки текстов.
- 13 Системное кодирование морфологической информации о служебных частях речи в системах автоматической обработки текстов.
- 14 Синтаксический словарь как компонент лингвистической базы знаний.
- 15 Синтаксическая модель управления глагола.
- 16 Семантический классификатор как компонент лингвистической базы знаний.
- 17 Семантическая классификация понятий, семантическая модель управления различных частей речи
- 18 Структура словарной статьи машинного словаря.
- 19 Словари ударений, сокращений, омонимов, синонимов, имен собственных, идиом в системах автоматической обработки языка. Разработка и использование.
- 20 Исходные понятия корпусной лингвистики: проблемная область, корпус данных, корпус текстов.
- 21 Параллельный многоязычный корпус текстов, его структура и сфера применения.
- 22 Требования к корпусу текстов с точки зрения пользователя.
- 23 Аннотированные корпуса текстов, автоматизация их создания и коррекции.
- 24 Практическое использование аннотированных корпусов текстов в системах автоматической обработки текстов.

- 25 Перевод как вид языковой деятельности. Общая формальная модель переводческой деятельности человека как основа ее моделирования в системах машинного перевода
- 26 Классификация переводов по различным признакам. Виды переводов в зависимости от степени автоматизации.
- 27 Машинный перевод как одно из направлений искусственного интеллекта. История возникновения и развития.
- 28 Подходы к моделированию процесса перевода в системах машинного перевода.
- 29 Оценка эффективности системы машинного перевода.
- 30 Основные компоненты системы машинного перевода.
- 31 Общий алгоритм задачи машинного перевода.
- 32 Лингвистические проблемы машинного перевода: проблемы на уровне машинного словаря.
- 33 Лингвистические проблемы машинного перевода: неточности перевода на уровне морфологии.
- 34 Лингвистические проблемы машинного перевода: неточности перевода на уровне синтаксиса.
- 35 Концептуальное поле омонимии. Системная классификация омонимов в устных и письменных текстах
- 36 Грамматическая омонимия как системная категория языка.
- 37 Пути возникновения омонимии. Межъязыковая омонимия
- 38 Способы снятия омонимии при автоматической обработке текстов. Алгоритм разрешения омонимии «глагол – имя существительное».
- 39 Способы снятия омонимии при автоматической обработке текстов. Алгоритм разрешения омонимии «именительный падеж – винительный падеж имени существительного».
- 40 Автоматический синтаксический анализ предложения. Основные этапы алгоритма синтаксического анализа.
- 41 Выделение именных групп в структуре предложения. Свойство проективности на уровне синтаксиса.
- 42 Алгоритм выделения простого предложения при автоматической обработке языка. Метод фильтров.
- 43 Автоматический семантический анализ предложения. Основные этапы.
- 44 Семантическая обработка слов, имеющих модель управления. Классификация предикатов и слов без модели управления.
- 45 Структурирование семантического графа. Группы кванторных слов.
- 46 Автоматическая обработка дискурса. Основные проблемы обработки связного текста. Интерпретация текста
- 47 Определение лингвистической взаимосвязи предложений дискурса. Установление соответствия референта и ссылки на него.
- 48 Восстановление эллипсиса при автоматической обработке дискурса.
- 49 Автоматический синтез предложений на естественном языке. Этапы лингвистического синтеза текста.
- 50 Автоматический синтез предложений на естественном языке. Задачи внелингвистического синтеза дискурса.

Критерии оценивания по экзамену

<i>Оценка</i>	<i>Критерии оценивания по экзамену</i>
<i>Высокий уровень «5» (отлично)</i>	<i>оценку «отлично» заслуживает студент, освоивший знания, умения, компетенции и теоретический материал без пробелов; выполнивший все задания, предусмотренные учебным планом на высоком качественном уровне; практические навыки профессионального применения освоенных знаний сформированы.</i>

<i>Средний уровень «4» (хорошо)</i>	<i>оценку «хорошо» заслуживает студент, практически полностью освоивший знания, умения, компетенции и теоретический материал, учебные задания не оценены максимальным числом баллов, в основном сформировал практические навыки.</i>
<i>Пороговый уровень «3» (удовлетворительно)</i>	<i>оценку «удовлетворительно» заслуживает студент, частично с пробелами освоивший знания, умения, компетенции и теоретический материал, многие учебные задания либо не выполнил, либо они оценены числом баллов близким к минимальному, некоторые практические навыки не сформированы.</i>
<i>Минимальный уровень «2» (неудовлетворительно)</i>	<i>оценку «неудовлетворительно» заслуживает студент, не освоивший знания, умения, компетенции и теоретический материал, учебные задания не выполнил, практические навыки не сформированы.</i>

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

- при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;
- при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;
- при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. Перечень учебной литературы, информационных ресурсов и технологий

5.1. Учебная литература

1. Боярский, К.К. Введение в компьютерную лингвистику [Электронный ресурс] : учебное пособие / К.К. Боярский. — Электрон. дан. — Санкт-Петербург : НИУ ИТМО, 2013. — 72 с. — Режим доступа: <https://e.lanbook.com/book/70822>. — Загл. с экрана.
2. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>.

5.2. Периодическая литература

1. Базы данных компании «Ист Вью» <http://dlib.eastview.com>

2. Электронная библиотека GREBENNICON.RU <https://grebennikon.ru/>

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» www.biblioclub.ru
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных:

1. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>

Ресурсы свободного доступа:

1. КиберЛенинка (<http://cyberleninka.ru/>);
2. Словари и энциклопедии <http://dic.academic.ru/>;

Собственные электронные образовательные и информационные ресурсы КубГУ:

1. Среда модульного динамического обучения <http://moodle.kubsu.ru>
2. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://mschool.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины (модуля)

Для успешного усвоения курса «Компьютерная лингвистика» и достижения поставленных целей и задач студенту необходимо:

- активно работать во время лабораторных занятий, участвуя во всех видах предлагаемой учебной деятельности (диалог с преподавателем в ходе занятия, выступления с рефератами и практическим анализом заданий, а также участие в групповой дискуссии);

- систематически самостоятельно работать в следующих направлениях:

- повторение базовых положений общего языкознания, истории языка и лексикологии, имеющих связь с курсом теоретической грамматики языка;

- составление глоссария базовых лингвистических и грамматических терминов и их толкований с транскрипцией и примерами;

- подготовка лексико-грамматических и синтаксических примеров анализа предложений, предлагаемых в рамках семинарских занятий, в письменной и устной форме;

- критическое изучение учебной и научной литературы по проблематике лабораторных занятий и её реферирование, сопоставление различных точек зрения и представление результатов в форме эссе;

- осуществление самоконтроля знаний основных теоретических положений теоретической грамматики и их применение в практике профессиональной деятельности.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного

контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

7. Материально-техническое обеспечение по дисциплине (модулю)

Наименование специальных помещений	Оснащенность специальных помещений	Перечень лицензионного программного обеспечения
Аудитория для проведения занятий лекционного типа, семинарских занятий, индивидуальных и групповых консультаций, промежуточной аттестации (305)	на 200 рабочих мест. учебная мебель; учебно-маркерная доска.	Мультимедийное оборудование: Проектор, Epson EB; ТВ, LG65 SM9010; интерактивная трибуна, TECOM TPRR-41; акустическая система show TA-S13M; камера HD Color Video Camera V1.0; учебная мебель; учебно-маркерная доска. Подключение к Интернету через беспроводную сеть Wi-Fi.
Компьютерный класс, Учебно-методическая лаборатория лингвистических исследований (321)	на 15 посадочных мест оснащена следующими техническими средствами обучения и оборудованием: учебная мебель; доска учебно-маркерная; проектор EB-585Wi; интерактивная доска (интерактивный проектор); обеспечено проводное подключение ПК к локальной сети и сети Интернет.	Офисное ПО: операционная система MS Windows, офисный пакет MS Office, платформа MS Teams.
Учебная аудитория для проведения индивидуальных и групповых консультаций (350)	на 15 посадочных мест оснащена следующими техническими средствами обучения и оборудованием: учебная мебель; учебно-маркерная доска	Мультимедийное оборудование: ПК-терминалы, 2 шт
Учебная аудитория для проведения индивидуальных и групповых консультаций (318)	на 5 посадочных мест оснащена следующими техническими средствами обучения и оборудованием: учебная мебель; доска учебно-маркерная; обеспечено проводное подключение ПК к локальной сети и сети Интернет.	Мультимедийное оборудование: ПК-терминалы, 2 шт
Кафедра прикладной лингвистики и новых информационных технологий, помещение для самостоятельной работы студентов (316)	учебная мебель; учебно-маркерная доска. Офисное ПО: операционная система MS Windows, офисный пакет MS Office, платформа MS Teams.	Мультимедийное оборудование: ПК-терминалы, 2 шт

Для самостоятельной работы обучающихся предусмотрены помещения, укомплектованные специализированной мебелью, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование помещений для самостоятельной работы обучающихся	Оснащенность помещений для самостоятельной работы обучающихся	Перечень лицензионного программного обеспечения
Помещение для самостоятельной работы обучающихся (читальный зал Научной библиотеки)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	Прикладное программное обеспечение (MicrosoftOffice)
Помещение для самостоятельной работы обучающихся (ауд. 347)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	Прикладное программное обеспечение (MicrosoftOffice)