

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталия Александровна
Должность: Ректор
Дата подписания: 24.10.2024 14:21:41
Уникальный программный ключ:
6b5279da4e034bff679172803da5b7b559fc69e2

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

УТВЕРЖДЕН
на заседании кафедры теории языка,
англистики и прикладной лингвистики

Протокол от «26» февраля 2024 г. №8

Зав. кафедрой Т.В. Холстинина 

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ТЕКУЩЕГО КОНТРОЛЯ**

по дисциплине

Компьютерная лингвистика

Направление подготовки:
45.03.02 Лингвистика

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

Квалификация
Бакалавр

Форма обучения
Очная

Мытищи
2024

Назначение

Осуществление текущего и промежуточного контроля по дисциплине «Компьютерная лингвистика»

Фонд оценочных средств текущего контроля разработан на основе рабочей программы дисциплины «Основы компьютерной лингвистики» в соответствии с требованиями Федерального Государственного образовательного стандарта высшего образования по направлению подготовки 45.03.02 Лингвистика, утвержденного приказом МИНОБРНАУКИ РОССИИ от 12.08.20, № 969

Разработчик:

Доктор филологических наук, профессор Максименко О.И.

СОДЕРЖАНИЕ

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы	4
2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания	4
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	7
4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций	11

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

Код и наименование компетенции	Этапы формирования
СПК-1 - способен использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач	1.Работа на учебных занятиях 2.Самостоятельная работа
СПК-2 - 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-3 - 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
СПК-4 - 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

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: основные понятия теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации Уметь: 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	Знать: основные понятия теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации Уметь: применять полученные знания для решения профессиональных задач Владеть: 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	Знать: Основные фонетические, лексические, грамматические, словообразовательные явления в языках мира; системы синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации Уметь: 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	Знать: Основные фонетические, лексические, грамматические, словообразовательные явления в языках мира; системы синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации Уметь: Применять полученные знания о структуре и системе языков мира в самостоятельных исследованиях Владеть:	Текущий контроль: Посещение занятий Выступление с докладом / сообщением Участие в групповом или индивидуальном проекте Тесты Задания для самостоятельного изучения	61–100 баллов

			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		
СПК-3	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знать: общенаучные методы и конкретные методики изучения данных в соответствующей области лингвистики; принципы работы с библиографическими источниками Уметь: использовать основные информационно-поисковые и экспертные системы, системы представления знаний в данной предметной области, принципы научно-доказательного изложения материала.	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	Знать: общенаучные методы и конкретные методики изучения данных в соответствующей области лингвистики; принципы работы с библиографическими источниками Уметь: использовать основные информационно-поисковые и экспертные системы, системы представления знаний в данной предметной области, принципы научно-доказательного изложения материала. Владеть: проблематикой смежных с лингвистикой областей и возможными подходами к их решению с позиций комплексного подхода	Текущий контроль: Посещение занятий Выступление с докладом / сообщением Участие в групповом или индивидуальном проекте Тесты Задания для самостоятельного изучения	61–100 баллов

СПК-4	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знать: математико-статистические методы обработки лингвистической информации, основы программирования, принципы автоматической обработки корпусов текстов Уметь: применять полученные знания для анализа и обработки нового лингвистического материала на изучаемых языках	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	Знать: математико-статистические методы обработки лингвистической информации, основы программирования, принципы автоматической обработки корпусов текстов Уметь: применять полученные знания для анализа и обработки нового лингвистического материала на изучаемых языках Владеть: способами представления полученных результатов, методикой изложения, принятой в соответствующей области лингвистического знания	Текущий контроль: Посещение занятий Выступление с докладом / сообщением Участие в групповом или индивидуальном проекте Тесты Задания для самостоятельного изучения	61–100 баллов

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

Тестовые задания

1. Что из перечисленного НЕ относится к уровням представления текста в лингвистическом процессоре:

- a) Фонетический
- b) Морфологический
- c) Синтаксический
- d) Семантический

2. Какие лексические элементы убирают из текста при информационном поиске и анализе данных:

- a) Имена собственные
- b) Слова с орфографическими ошибками
- c) Стоп-слова
- d) Шумовые выбросы

3. Линейная прогрессия, при которой одна задача выполняется после другой в определенной последовательности, это:

- a) Последовательность (sequence)
- b) Зацикливание (while)
- c) Функция выбора (if-then-else)

4. Функция, при которой делается выбор между двумя вариантами действий:

- a) Последовательность (sequence)
- b) Зацикливание (while)
- c) Функция выбора (if-then-else)

5. Элементы синтаксиса, которые связывают несколько инструкций в одну:

- a) Строка
- b) Список
- c) Блок
- d) Класс

6. Как разрешить многозначность лексических единиц на уровне конечных автоматов?

- a) При помощи преобразования конечного автомата в конечных преобразователь
- b) При помощи составления каскада конечных автоматов
- c) При помощи добавления блока правил
- d) На уровне конечных автоматов невозможно проводить дизамбигуацию

7. Описание структуры алгоритма на естественном языке называется:

- a) Словесным
- b) Формальным
- c) Блок-схемой
- d) Графиком

8. Описание структуры алгоритма на языке программирования называется:

- a) Словесным
- b) Формальным
- c) Блок-схемой
- d) Графиком

9. Графическое представление метода решения задачи называется:

- a) Словесным описанием
- b) Формальным описанием
- c) Блок-схемой
- d) Графиком

10. Алгоритм, в котором последовательность выполнения операций зависит от

определенных условий, называется:

- a) Линейным
- b) Разветвляющим
- c) Циклическим
- d) Рекурсивным

Ключ

№ вопр	Правильный ответ
1	b
2	a
3	c
4	a
5	b
6	b
7	c
8	b
9	c
10	a

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

Темы групповых проектов/сообщений

1. Составить алгоритм токенизации.
2. Сравнение подходов к описанию алгоритма у разных исследователей.
3. Формальное описание алгоритма извлечения данных из 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.

Промежуточная аттестация.

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

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.

Рекомендуемая тематика 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. Методы разрешения анафоры
20. Методы частеречной разметки для русского языка
21. Описание морфологических правил русского языка при помощи конечных преобразователей
22. Основные подходы к автоматическому синтаксическому анализу
23. Подходы к дизамбигуации
24. Подходы к извлечению именованных сущностей
25. Преобработка текста
26. Применение конечных автоматов в лингвистических задачах
27. Применение регулярных выражений для извлечения данных
28. Принципы создания контролируемого языка
29. Псевдокод как инструмент описания алгоритмов
30. Сравнительный анализ информационно-поисковых систем
31. Сравнительный анализ морфологических анализаторов для английского языка
32. Сравнительный анализ морфологических анализаторов для русского языка
33. Сравнительный анализ систем распознавания речи
34. Сравнительный анализ систем синтаксического анализа
35. Сравнительный анализ систем синтеза речи
36. Структура гипертекста

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

Этап: проведение текущего контроля успеваемости по дисциплине

Вид работы	Пятибалльная шкала	Шкала оценивания
1. Посещение занятий (лекции, практические занятия)	«отлично»	10 баллов , если студент посетил 90% от всех занятий
	«хорошо»	8 баллов , если студент посетил как минимум 70% от всех занятий
	«удовлетворительно»	5 баллов если студент посетил как минимум 50% от всех занятий
	«неудовлетворительно»	3 балла , если из всех занятий студент посетил как минимум 25%
2. Ответы на практическом занятии	«отлично»	30 баллов ответ полный, с привлечением знаний из разных разделов курса; 5 и

		более примеров
	«хорошо»	20 баллов ответ полный; 3-4 примера
	«удовлетворительно»	15 баллов ответ неполный; 1-2 примера
	«неудовлетворительно»	13 баллов ответ не соответствует теоретическому вопросу; без примеров
3. Тестирование	«отлично»	25 баллов , если из всех заданий студент выполнил как минимум 70%
	«хорошо»	22 баллов , если из всех заданий студент выполнил как минимум 60%
	«удовлетворительно»	17 баллов , если из всех заданий студент выполнил 40%-59%
	«неудовлетворительно»	9 баллов , если из всех заданий студент выполнил 20%
4. Самостоятельная подготовка к практическим занятиям	«отлично»	25 баллов , если из всех заданий студент выполнил как минимум 80%
	«хорошо»	22 баллов , если из всех заданий студент выполнил как минимум 60%
	«удовлетворительно»	17 баллов , если из всех заданий студент выполнил 40%
	«неудовлетворительно»	11 баллов , если из всех заданий студент выполнил 20%
5. Зачет с оценкой	«отлично»	10 баллов
	«хорошо»	8 баллов
	«удовлетворительно»	6 баллов
	«неудовлетворительно»	4 балла

Промежуточная аттестация

Шкала оценивания ответа на зачете с оценкой

Уровень овладения	неудовлетворительный	удовлетворительный	оптимальный	высокий
Дескрипторы				

Полнота ответа на теоретический вопрос	4 Ответ, 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 практическим и 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 (максимальный)	18	25	32	40
	не зачтено	зачтено	зачтено	зачтено

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	10	7	6	5

изучения				
Зачет	40	35	28	20
Итого	81- 100	61-80	41-60	0-40

- оценка «отлично» выставляется студенту, набравшему 81–100 баллов;
- оценка «хорошо» выставляется студенту, набравшему 61–80 баллов;
- оценка «удовлетворительно» выставляется студенту, набравшему 41–60 баллов;
- оценка «неудовлетворительно» выставляется студенту, набравшему 40–0 баллов.

Шкала оценивая курсовой работы

1. Курсовая работа	«отлично»	100 баллов , 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.
	«хорошо»	80 баллов , если представленный на защите продукт свидетельствует о проведенном исследовании с привлечением различных источников информации, отражает теоретическую и практическую направленность проекта.
	«удовлетворительно»	60 баллов , если представленный на защите продукт свидетельствует о проведенном исследовании, отражает теоретическую направленность проекта.
	«неудовлетворительно»	40 балла , если представленный на защите продукт свидетельствует о проведенном исследовании с привлечением одного источника информации; слабо отражает теоретическую и практическую направленность проекта.