

Документ подписан электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 14.05.2024 08:54:15
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
ФГБОУ ВО «ТВЕРСКОЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Утверждаю:
Руководитель ООП
Л.М. Сапожникова


«5» мая 2024 г.


Рабочая программа дисциплины

Информационные технологии в лингвистике

Закреплена за кафедрой: **Герменевтической лингводидактики и английской филологии**

Направление подготовки: **45.03.02 Лингвистика**

Направленность (профиль): **Лингвистика, перевод и межкультурная коммуникация**

Квалификация: **Бакалавр**

Форма обучения: **очная**

Семестр: **2**

Программу составил(и):
канд. филол. наук, доц., Колосов Сергей Александрович; канд. филол. наук, доц., Масленникова Евгения Михайловна

Тверь, 2024

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цели освоения дисциплины (модуля):

Формирование и развитие у обучающихся профессиональных компетенций, связанных с информационно-лингвистической деятельностью, а именно: навыков работы с компьютером как средством получения, обработки и управления информацией; умения работать с традиционными носителями информации, распределёнными базами данных и знаний; способности работать с информацией в глобальных компьютерных сетях, с электронными словарями и другими электронными ресурсами для решения общепрофессиональных задач.

Задачи:

- овладение понятийным аппаратом теоретической и прикладной лингвистики в области разработки и применения информационных технологий;
- овладение основами информационной и библиографической культуры;
- развитие умений использовать достижения отечественного и зарубежного научного наследия, современных прикладных направлений лингвистики для решения конкретных методических и переводческих задач;
- знакомство с областью применения и функционалом электронных лингвистических ресурсов
- формирование умения формализовать языковой материал в соответствии с поставленными задачами;
- овладение стандартными методиками поиска, анализа и обработки данных, необходимых в профессиональной деятельности.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ОП: Б1.О

Требования к предварительной подготовке обучающегося:

Иностранный язык

Введение в языкознание

Общее языкознание

Практический курс первого иностранного языка

Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:

Электронные ресурсы переводчика

Практика письменного перевода

Введение в методологию научных исследований

3. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость	2 ЗЕТ
Часов по учебному плану	72
в том числе:	
аудиторные занятия	36
самостоятельная работа	36

4. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

ОПК-5.1: Использует основные методы поиска, сбора, обработки и систематизации информации в лингвистике; современные программные средства редактирования текстовых документов; нормативно закрепленные требования к оформлению текста

ОПК-5.2: Осуществляет поиск, сбор информации в компьютерной сети

ОПК-5.3: Проводит обработку и систематизацию информации; редактирует и оформляет тексты на компьютере в различных текстовых редакторах

ОПК-6.1: Интерпретирует основные положения корпусной лингвистики

ОПК-6.2: Осуществляет поиск и сбор информации в лингвистических корпусах для решения профессиональных задач

ОПК-6.3: Использует лингвистические информационные ресурсы для решения конкретных методических и переводческих задач

УК-1.1: Анализирует задачу, выделяя ее базовые составляющие

УК-1.2: Определяет, интерпретирует и ранжирует информацию, требуемую для решения поставленной задачи;

УК-1.3: Осуществляет поиск информации для решения поставленной задачи по различным типам запросов

УК-1.4: При обработке информации отличает факты от мнений, интерпретаций, оценок, формирует собственные мнения и суждения, аргументирует свои выводы и точку зрения

УК-1.5: Рассматривает и предлагает возможные варианты решения поставленной задачи, оценивая их достоинства и недостатки

5. ВИДЫ КОНТРОЛЯ

Виды контроля в семестрах:	
зачеты	2

6. ЯЗЫК ПРЕПОДАВАНИЯ

Язык преподавания: русский.

7. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занят.	Наименование разделов и тем	Вид занятия	Семестр / Курс	Часов	Источники	Примечание
	Раздел 1. Тематическое содержание дисциплины					
1.1	Введение в предметную область.	Лек	2	2	Л1.1 Л1.3	
1.2	Основные понятия информационного поиска и библиографического описания.	Лек	2	4	Л1.1 Л1.3	
1.3	Современные программные средства работы с текстовыми документами и нормативно закрепленные требования к оформлению текста.	Лек	2	2		
1.4	Электронная лексикография.	Лек	2	2		
1.5	Автоматический анализ и синтез текста.	Лек	2	2		
1.6	Корпусная лингвистика.	Лек	2	2	Л1.2	
1.7	Машинный перевод и автоматизированный перевод.	Лек	2	4	Л1.1	

1.8	Основы работы в ЭОС ТвГУ LMS Canvas. Создание портфолио в ЭОС ТвГУ LMS Canvas.	Пр	2	4		
1.9	Основы поиска в сети Интернет. Поисковые системы.	Пр	2	2	Л1.3	
1.10	Работа в текстовом редакторе Word. Редактирование текста в соответствии с нормативными требованиями.	Пр	2	4		
1.11	Лингвистические порталы.	Пр	2	2		
1.12	Электронные словари.	Пр	2	2		
1.13	Электронные корпуса. Поиск сочетаемости слов с помощью электронных корпусов.	Пр	2	2		
1.14	Работа с сервисами Google: Основы работы в системе Google Диск.	Пр	2	2		
	Раздел 2. Самостоятельная работа					
2.1	Основные понятия информационного поиска и библиографического описания.	Ср	2	4		
2.2	Современные программные средства работы с текстовыми документами и нормативно закрепленные требования к оформлению текста.	Ср	2	8		
2.3	Основы работы в ЭОС ТвГУ LMS Canvas. Создание портфолио в ЭОС ТвГУ LMS Canvas.	Ср	2	2		
2.4	Электронная лексикография.	Ср	2	4		
2.5	Корпусная лингвистика.	Ср	2	8		
2.6	Лингвистические порталы.	Ср	2	4		
2.7	Работа с сервисами Google: Основы работы в системе Google Диск.	Ср	2	4		
2.8	Машинный перевод и автоматизированный перевод.	Ср	2	2		

Список образовательных технологий

1	Дискуссионные технологии (форум, симпозиум, дебаты, аквариумная дискуссия, панельная дискуссия, круглый стол, фасилитированная и т.д.)
2	Информационные (цифровые) технологии

3	Активное слушание
4	Метод case-study
5	Тренинг
6	Портфолио

8. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

8.1. Оценочные материалы для проведения текущей аттестации

Примерные образцы заданий для проведения текущей аттестации

Задание 1.

С помощью электронного этимологического словаря М. Фасмера установите этимологию слова autumn в английском. Представьте результаты поиска в виде отредактированного документа в текстовом редакторе в соответствии с требованиями (выбор шрифта, кегля, выделения текста, выравнивания, междустрочного интервала, расстановка переносов и т.д.). Вставьте скан страницы со словарной дефиницией.

Задание 2.

С помощью электронного этимологического словаря М. Фасмера установите этимологию слова осень в русском языке. Представьте результаты поиска в виде отредактированного документа в текстовом редакторе в соответствии с требованиями (выбор шрифта, кегля, выделения текста, выравнивания, междустрочного интервала, расстановка переносов и т.д.). Вставьте скан страницы со словарной дефиницией.

Задание 3.

С помощью программы «Visual Thesaurus» (<https://www.visualthesaurus.com/>) установить и построить семантические сети для существительных autumn и fall в английском языке. Представить результаты поиска в виде отредактированного документа в текстовом редакторе в соответствии с требованиями оформления согласно требованиям ГОСТа (выбор шрифта, кегля, выделения текста, выравнивания, междустрочного интервала, расстановка переносов и т.д.). Вставить схемы (рисунки).

Задание 4. Тест.

1. Область языкознания, связанная с разработкой методов решения практических задач использования языка и способов оптимизации функционирования языка называется _____ лингвистика.

- А) фундаментальная
- Б) прикладная
- В) количественная
- Г) компьютерная

2. Раздел лингвистики, исследующий язык и его функционирование при помощи статистических методов, называется _____ лингвистика.

- А) прикладная
- Б) компьютерная
- В) количественная
- Г) эмпирическая

3. Релевантность, ценность и актуальность относятся к _____ свойствам информации.

- А) сущностным
- Б) семантическим
- В) прагматическим

4. Релевантность результатов поиска определяется _____.
- А) соответствием выданных документов поисковому запросу
 - Б) количеством выданных документов
 - В) количеством документов, имеющихся в системе
 - Г) всеми выше указанными факторами
5. Множество документов, выдаваемых в процессе информационного поиска, формально соответствующих запросу (релевантных), но не являющихся релевантными по смыслу, называется _____.
- А) базой данных
 - Б) полнотой поиска
 - В) информационным шумом
 - Г) информационно-поисковым тезаурусом
6. В результате поиска объём выдачи составил 150 документов, при этом число релевантных документов оказалось равным 60. Показатель точности поиска равняется _____.
- А) 25%
 - Б) 40%
 - В) 60%
 - Г) 90%
7. Для открытия окна поиска на веб-странице или в текстовом документе используется сочетание клавиш _____.
- А) Alt + D
 - Б) Ctrl + C
 - В) Ctrl + F
8. Раздел прикладной лингвистики, занимающийся разработкой и использованием электронных словарей, называется _____.
- А) компьютерная лингвистика
 - Б) компьютерная лексикография
 - В) корпусная лингвистика
 - Г) машинный перевод
9. Словарь, в котором словарные единицы сгруппированы по семантическим полям и понятийным группам с указанием семантических отношений между единицами, называется _____.
- А) конкорданс
 - Б) тезаурус
 - В) энциклопедия
 - Г) толковый словарь
10. Термин «компьютерная лингвистика» по-английски звучит как _____ linguistics. (необходимо ввести слово)
11. Связь между языковыми явлениями, при которой одно из явлений входит в число причин, определяющих другие, или когда имеются общие причины, воздействующие на эти явления, называется _____.
- А) модель данных
 - Б) корреляция
 - В) дихотомия
 - Г) диахрония
12. Формат KWIC при построении конкорданса означает, что _____.

- А) список примеров выровнен по ключевому слову в центре строки
- Б) список примеров представлен в алфавитном порядке
- В) представлены только самые частотные сочетания с ключевым словом

13. Главной характеристикой, которая отличает лингвистический корпус от коллекций (или «библиотек») текстов и других лингвистических баз данных, является _____.

- А) разметка
- Б) тип данных
- В) репрезентативность

14. «Корпус содержит по возможности все типы письменных и устных текстов, представленные в языке, и все эти тексты входят в корпус по возможности пропорционально их доле в языке соответствующего периода». Данное описание характеризует такой параметр корпуса, как

- А) аннотирование
- Б) структурированность
- В) репрезентативность
- Г) диахронность

15. Прикладная дисциплина, в центре внимания которой стоят принципы и методы создания словарей, называется _____. (необходимо ввести слово)

16. Тезаурус представляет собой _____.

- А) толковый словарь
- Б) идеографический словарь
- В) словарь сочетаемости
- Г) многоязычный словарь

17. Что из нижеперечисленного относится к системам контекстуального поиска по переводам?

- А) DeepL
- Б) Trados
- В) Linguee
- Г) Каргаслов.ру

18. Утверждение «автоматизированный перевод это то же самое что машинный перевод»

- А) верно
- Б) неверно

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

- А) база данных
- Б) тезаурус
- В) корпус
- Г) гипертекст

20. Список всех контекстов употребления искомой единицы в корпусе называется _____.

- А) конкорданс
- Б) конкордансер
- В) тезаурус
- Г) коллокация

21. Корпусная лингвистическая разметка текстов НЕ бывает _____.
- А) морфологической
 - Б) синтаксической
 - В) просодической
 - Г) диахронической
22. Программа, осуществляющая морфологическую или синтаксическую разметку текста в корпусе, называется _____.
- А) парсер
 - Б) корпусный менеджер
 - В) конкордансер
 - Г) токен
23. В морфологическую разметку корпуса НЕ входит _____.
- А) частеречная принадлежность
 - Б) признаки грамматических категорий
 - В) форматирование текста
24. Элементы разметки корпуса, передающие внешнюю, экстралингвистическую информацию о тексте (имя автора, год создания, жанр, тематика и т.д.), относятся к _____.
- А) семантической разметке
 - Б) структурной разметке
 - В) метаразметке
25. Два слова, которые в тексте или корпусе текстов являются соседними, называются _____.
- А) конкорданс
 - Б) коллокация
 - В) биграмма
 - Г) дихотомия
26. Английская аббревиатура CAT обозначает _____.
- А) перевод, выполняемый человеком с использованием технологии памяти переводов
 - Б) перевод, полностью выполняемый машиной (компьютером)
 - В) перевод, выполненный в специальном текстовом редакторе
27. Предложение Математическая модель перевода описывается рядом формул содержит _____ триграмм(ы).
- А) 2
 - Б) 3
 - В) 4
 - Г) 6
28. В основе системы статистического машинного перевода лежит использование _____.
- А) параллельных корпусов текстов
 - Б) корпуса электронных словарей
 - В) формальной грамматики, описанной с помощью математических формул
 - Г) заданного алгоритма перевода
29. Достаточно высокая синтаксическая и морфологическая точность, а также стабильность и предсказуемость результата являются преимуществами _____.
- А) систем машинного перевода, основанного на правилах
 - Б) систем статистического машинного перевода

30. Системы Trados, Memsource и Wordfast относятся к системам _____.

- А) машинного перевода
- Б) информационного поиска
- В) памяти переводов
- Г) автоматического синтеза текста

31. Программа, отвечающая на вопросы пользователя на корпоративной веб-странице в автоматическом режиме, называется _____.

- А) парсер
- Б) корпусный менеджер
- В) бот
- Г) мессенджер

32. Процесс преобразования лексемно-морфологических, синтаксических и семантических представлений в текст на естественном языке называется _____.

- А) автоматическим синтезом текста
- Б) автоматическим анализом текста
- В) автоматическим переводом

33. Типовая парадигма лексемы в автоматическом морфологическом анализе – это _____.

- А) последовательность букв от начала словоформы, общая для всех словоформ
- Б) элементы, описывающие формоизменение конкретной лексемы
- В) совокупность наборов машинных окончаний
- Г) совпадение основ разных слов

34. Полная морфологическая характеристика Деепричастия включает набор следующих грамматических признаков:

- А) вид, переходность
- Б) вид, переходность, время
- В) вид, переходность, число, род

35. Полная морфологическая характеристика Имени прилагательного в русском языке включает набор следующих грамматических признаков:

- А) род, число, падеж
- Б) род, число, падеж, форма, степень сравнения
- В) род, число, падеж, форма, степень сравнения, переходность
- Г) род, число, падеж, форма, степень сравнения, разряд

36. Полная морфологическая характеристика Глагола в русском языке включает набор следующих грамматических признаков:

- А) вид, время, число
- Б) наклонение, вид, переходность, время
- В) наклонение, вид, переходность, время, число, род
- Г) наклонение, вид, залог, переходность, время

37. Расположите этапы автоматического анализа текста в порядке их выполнения:

- А) морфологический анализ
- Б) синтаксический анализ
- В) графематический анализ
- Г) семантический анализ

38. Выберите правильную последовательность этапов автоматического синтеза текста:

- А) семантический, синтаксический, морфологический, графематический

- Б) графематический, морфологический, синтаксический, семантический
В) семантический, морфологический, синтаксический, графематический

39. В каком предложении имеется синтаксическая омонимия?

- А) Андрей вчера сдавал экзамен по математике.
Б) Нина читала книгу матери.
В) Расскажите о своей семье.
Г) Маша читает книгу о животных.

40. В каком из предложений НЕТ синтаксической омонимии?

- А) Родные тётки ждали дядю Игоря.
Б) Аня поставила на полку новые учебники и словари.
В) Преподаватель вернулся из командировки в Москву.
Г) Он надел пальто на улице.

41. Какая словоформа в предложении Девочка мыла окно представляет собой случай лексико-морфологической омонимии?

- А) девочка
Б) мыла
В) окно
Г) никакая

42. При построении дерева зависимостей исходным пунктом синтаксического анализа является _____.

- А) подлежащее
Б) сказуемое
В) именная группа
Г) предложение

43. Начальная форма слова, к которой можно привести все отдельные словоупотребления и грамматические формы одной лексемы, называется

- А) токен
Б) парсер
В) лемма

44. Страницы Википедии на разных языках, посвящённые одной теме, являются примером _____ корпуса.

- А) параллельного
Б) сравнительного
В) выровненного
Г) смешанного

45. В каком году состоялся Джорджтаунский эксперимент?

- А) 1947
Б) 1954
В) 1979
Г) 1983

8.2. Оценочные материалы для проведения промежуточной аттестации

1. Итоговый тест по лекционным материалам курса (в LMS).
2. Комплексное практическое задание.

Образец комплексного задания.

С помощью программы «Visual Thesaurus» (<https://www.visualthesaurus.com/>) установить и построить семантические сети для существительных autumn и fall в

английском языке. Представить результаты поиска в виде отредактированного документа в текстовом редакторе в соответствии с требованиями оформления согласно требованиям ГОСТа (выбор шрифта, кегля, выделения текста, выравнивания, междустрочного интервала, расстановка переносов и т.д.). Вставить схемы (рисунки).

8.3. Требования к рейтинг-контролю

Максимальное количество баллов – 100, в том числе

20 баллов – выполнение итогового теста по материалам лекционного курса

80 баллов – выполнение практических заданий и работа в аудитории.

9. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

9.1. Рекомендуемая литература

9.1.1. Основная литература

Шифр	Литература
Л1.1	Шунейко, Авдеенко, Квантитативная лингвистика и новые информационные технологии, Москва: Юрайт, 2024, ISBN: 978-5-534-15446-7, URL: https://urait.ru/bcode/543983
Л1.2	Шунейко, Корпусная лингвистика, Москва: Юрайт, 2024, ISBN: 978-5-534-13603-6, URL: https://urait.ru/bcode/543746
Л1.3	Гаврилов, Климов, Информатика и информационные технологии, Москва: Юрайт, 2024, ISBN: 978-5-534-15819-9, URL: https://urait.ru/bcode/535560

9.3.1 Перечень программного обеспечения

1	Google Chrome
2	ABBYY Lingvo x5
3	OpenOffice
4	Многофункциональный редактор ONLYOFFICE

9.3.2 Современные профессиональные базы данных и информационные справочные системы

1	ЭБС «ZNANIUM.COM»
2	ЭБС «ЮРАИТ»
3	ЭБС «Университетская библиотека онлайн»
4	ЭБС IPRbooks
5	ЭБС «Лань»
6	ЭБС ТьГУ

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Аудит-я	Оборудование
2-405	комплект учебной мебели, дубликатор, компьютеры, листоподборщик, фальцовщик, коммутатор
2-407	комплект учебной мебели, VHS/DVD, компьютеры, акустические системы для фонового озвучивания, дека, доска флипчарт, усилитель
2-206	комплект учебной мебели, компьютер, мультимедийный комплект учебного класса (вариант № 2), проектор, переносной ноутбук, экран настенный, тумба

11. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

План лекционных занятий (теоретический блок)

Тема 1. Введение в предметную область

- Информация и ее свойства.
- Информационные технологии.
- Лингвистика как наука.
- Фундаментальная и прикладная лингвистика.
- Компьютерная лингвистика.
- Квантитативная лингвистика.
- Основные направления прикладной и компьютерной лингвистики.

Тема 2. Основные понятия информационного поиска

- Типы поисковых систем.
- Принципы работы поисковых систем.
- Релевантность, полнота и точность поиска.
- Принципы составления поисковых запросов и оптимизация поиска в

интернете.

Тема 3. Современные программные средства работы с текстовыми документами и нормативно закреплённые требования к оформлению текста.

- Текстовый редактор Word.
- Горячие клавиши.
- Требования к оформлению ВКР.

Тема 4. Информационные лингвистические ресурсы

- Типы лингвистических ресурсов.
- Информационные ресурсы в лингвистике.
- Электронные библиотеки.
- Основы библиографической культуры.

Тема 5. Электронная лексикография

• Типы словарей.
• Недостатки традиционных бумажных словарей и преимущества электронного словаря как особого лексикографического объекта.

- Словари тезаурусного типа.
- Электронные одноязычные и многоязычные словари.
- Особенности электронного словаря, существующего в среде многомерного

гипертекста.

- Системы контекстуального поиска по переводам.

Тема 6. Корпусная лингвистика и лингвистические корпуса

- Прикладные аспекты корпусной лингвистики.
- Требования к лингвистическим корпусам.
- Типы лингвистических корпусов.
- Виды корпусной разметки.
- Корпусные менеджеры и поисковые запросы.
- Конкордансы.

Тема 7. Автоматический анализ и синтез звучащей речи. Автоматический анализ и синтез текста

• Понятия автоматического анализа и синтеза речи и текста: проблемы и сферы применения.

• Графематический, морфологический, синтаксический и семантический анализ текста.

- Понятия токенизации, лемматизации, парсинга.
- Формальная грамматика.
- Автоматический синтез текста.

Тема 8. Машинный перевод и автоматизированный перевод

- Краткая история машинного перевода.
- Машинный перевод, основанный на правилах.

- Статистический машинный перевод.
- Нейронный машинный перевод.
- Системы автоматизированного перевода (CAT-инструменты): принципы работы, наиболее распространённые программы.

Темы для рефератов:

- Цели и задачи использования информационных технологий
- Понятие «прикладная лингвистика»
- Основные направления прикладной лингвистики
- Виды электронных ресурсов
- Виды гипертекстов (отличие электронного гипертекста от бумажного)
- Особенности словарей разных видов
- Особенности электронных библиотек
- Особенности электронных словарей
- Мультимедийные энциклопедии
- Понятие «конкорданс» и «конкордансер»
- Понятие «тезаурус». Виды тезаурусов
- Цели и задачи корпусной лингвистики
- Корпуса немецкого языка
- Национальный корпус русского языка
- Многоязычные корпуса
- Типы корпусной разметки
- Возможности использования корпусов национальных языков для решения научно-исследовательских задач
- Возможности использования корпусов национальных языков при обучении языкам