

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 27.09.2023 09:05:19
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Тверской государственный университет»

Утверждаю:

Руководитель ООП

_____ Л. М. Сапожникова

«__» _____ 2021 г.

Рабочая программа дисциплины (с аннотацией)
«Информационные технологии в лингвистике»

Направление подготовки

45.03.02 «Лингвистика»

Профили подготовки

*«Перевод и переводоведение»,
«Теория и методика преподавания иностранных языков и культур»*

Для студентов 1 курса очной формы обучения
и 2 курса заочной формы обучения

БАКАЛАВРИАТ

Составители: *к.ф.н. С. А. Колосов,
к.ф.н. С. Ю. Воскресенская
к.ф.н. К. А. Гудий
к.ф.н. Е. М. Масленникова*

Тверь, 2021

I. Аннотация

1. Наименование дисциплины (или модуля) в соответствии с учебным планом

Информационные технологии в лингвистике

2. Цели и задачи дисциплины (модуля)

Целью освоения дисциплины является 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.

Задачами освоения дисциплины являются:

- 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.

3. Место дисциплины в структуре ООП бакалавриата

Учебная дисциплина «Информационные технологии в лингвистике» входит в базовую часть учебного плана направления 45.03.02 Лингвистика и формирует ОПК у всех профилей. Вместе с дисциплинами «Информатика», «Электронные ресурсы преподавателя и переводчика» дисциплина «Информационные технологии в лингвистике» призвана расширять и углублять знания студентов в области новых информационных технологий и на этой базе развивать умения будущих лингвистов, преподавателей и переводчиков проводить экспертизу лингвистических программных продуктов и использовать их в дальнейшей учебной и научно-исследовательской деятельности. Успешное освоение дисциплины «Информационные технологии в лингвистике» обеспечивает подготовку к осуществлению информационно-лингвистической деятельности и научно-исследовательской деятельности.

Освоение дисциплины «Информационные технологии в лингвистике» необходимо для последующих курсов по дисциплинам: «Электронные ресурсы преподавателя и переводчика», «Современные научные исследования в лингвистике», «Профессиональная социализация со знанием ИЯ», «Введение в методологию научных исследований», «Методика преподавания ИЯ», для производственной практики в соответствии с профилем, а также для подготовки выпускной квалификационной работы с использованием информационных технологий.

4. Объём дисциплины (или модуля):

Для очной формы обучения:

___3___ зачётных единицы, ___108___ академических часов, в том числе

контактная работа: лекции ___19___ часов, практические занятия _____ часов, лабораторные работы ___19___ часов, **самостоятельная работа:** ___70___ часов.

Для заочной формы обучения:

___3___ зачётных единицы, ___108___ академических часов, в том числе

контактная работа: лекции ___6___ часов, практические занятия _____ часов, лабораторные работы ___4___ часов, **самостоятельная работа:** ___89___ часов и контроль _____9_____ часов.

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

Бакалавр по направлению подготовки 45.03.02 «Лингвистика» должен продемонстрировать необходимый уровень сформированности **общепрофессиональных компетенций**, в соответствии с которыми он способен:

- работать с электронными словарями и другими электронными ресурсами для решения лингвистических задач (ОПК-13);
- владеть основами современной информационной и библиографической культуры (ОПК-14);
- уметь решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учётом основных требований информационной безопасности (ОПК-20).

Планируемые результаты освоения образовательной	Планируемые результаты обучения по дисциплине (или модулю)
--------------------------------------------------------	-------------------------------------------------------------------

программы (формируемые компетенции)	
<p>ОПК-13 способность работать с электронными словарями и другими электронными ресурсами для решения лингвистических задач</p>	<p>Владеть: навыками работы с электронными лингвистическими ресурсами;</p> <p>Уметь: решать профессиональные лингвистические задачи с помощью различных электронных ресурсов (словари, корпуса текстов);</p> <p>Знать: основные типы электронных лингвистических ресурсов (словари, корпуса текстов) и их функциональные возможности для решения лингвистических задач.</p>
<p>ОПК-14 владение основами современной информационной и библиографической культуры</p>	<p>Владеть: навыками информационного поиска в сети Интернет; нормами оформления библиографических данных;</p> <p>Уметь: редактировать и форматировать текст в компьютерном текстовом редакторе в соответствии с поставленной задачей; формировать поисковый запрос;</p> <p>Знать: основные алгоритмы информационного поиска и базовые операторы языковых запросов; требования к оформлению библиографической информации.</p>
<p>ОПК-20 способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности</p>	<p>Владеть: навыками решения стандартных профессиональных задач с применением информационно-лингвистических технологий и электронных ресурсов;</p> <p>Уметь: формализовать языковой материал для решения поставленной задачи; проводить отбор и анализ языковых фактов с целью применения информационных технологий;</p> <p>Знать: основные области применения информационных технологий в лингвистике; терминологический аппарат, связанный с использованием информационных технологий в лингвистике; основные составляющие информационно-коммуникационной компетенции лингвиста.</p>

6. Форма промежуточной аттестации

экзамен (2 семестр). Для заочной формы обучения – экзамен в зимнюю сессию.

7. Язык преподавания русский.

II. Содержание дисциплины (или модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

1. Для студентов очной формы обучения:

Наименование разделов и тем	Всего	Контактная работа (час.)		Самостоятельная работа
		Лекции	Практические работы	
Основные понятия курса «Информационные технологии в лингвистике».		2		4
Создание и использование баз данных.		1	4	8
Основы информационного поиска и библиографической культуры.		3	4	8
Автоматический анализ и синтез звучащей речи; автоматический анализ и синтез текста.		4		4
Машинный перевод и автоматизированный перевод.		4		4
Электронная лексикография.		2	4	8
Корпусная лингвистика.		3	4	8
Работа с текстовым редактором.			3	6
Подготовка реферата и подготовка к экзамену				20
ИТОГО	108	19	19	70

2. Для студентов заочной формы обучения:

Наименование разделов и тем	Всего	Контактная работа (час.)		Самостоятельная работа
		Лекции	Практические работы	
Основные понятия курса «Информационные технологии в лингвистике».		1		10
Создание и использование баз данных.		1	1	10
Основы информационного поиска и библиографической культуры.		1	1	10
Автоматический анализ и синтез звучащей речи; автоматический анализ и синтез текста.		1		10
Машинный перевод и автоматизированный перевод.		1		10
Электронная лексикография.		1		10

Корпусная лингвистика.			1	15
Работа с текстовым редактором.			1	14
Контроль				9
ИТОГО	108	6	4	89+9

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

- электронные презентации в Power Point для сопровождения лекционного курса
- программа каждой учебной темы
- глоссарий основных терминов курса
- список контрольных вопросов по содержанию курса для текущего и рубежного контроля
- банк типовых заданий для практических занятий
- тематика рефератов
- список экзаменационных вопросов по дисциплине.

IV. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине (или модулю)

1. **Оценочное средство** для проверки компонента «знать» формируемых компетенций: **ОПК-13** (способность работать с электронными словарями и другими электронными ресурсами для решения лингвистических задач); **ОПК-14** (владение основами современной информационной и библиографической культуры); **ОПК-20** (способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности).

Задание выполняется в виде теста с закрытыми типами вопросов.

ПРИМЕР ТЕСТА

1. Раздел прикладной лингвистики, занимающийся разработкой и использованием электронных словарей, называется

- А) компьютерная лингвистика
- Б) компьютерная лексикография
- В) корпусная лингвистика
- Г) машинный перевод

2. Словарь, в котором словарные единицы сгруппированы по семантическим полям и понятийным группам с указанием семантических отношений между единицами, называется _____.

- А) конкордансом
- Б) тезаурусом
- В) энциклопедией
- Г) толковым

3. Файлы-изображения и звуковые файлы в табличных базах данных содержатся в _____.

- А) поле «гиперссылка»
- Б) поле «логические данные»
- В) поле «объект OLE»

Г) поле «текстовые данные»

4. 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

5. Программа, осуществляющая морфологическую или синтаксическую разметку текста в корпусе, называется _____.

- А) парсер
- Б) корпусный менеджер
- В) конкордансер
- Г) токен

6. 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

7. Системы Trados и Wordfast относятся к системам _____.

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

2. Типовые контрольные задания для проверки уровня сформированности компетенции 1 (ОПК-13 – способность работать с электронными словарями и другими электронными ресурсами для решения лингвистических задач)

Этап формирования компетенции, в котором участвует дисциплина	Типовые контрольные задания для оценки знаний, умений, навыков (2-3 примера)	Показатели и критерии оценивания компетенции, шкала оценивания
владеть навыками работы с электронными словарями и другими электронными ресурсами для решения лингвистических задач <i>Базовый</i>	<i>Написание реферата</i>	<i>Полнота раскрытия темы</i> <ul style="list-style-type: none">• Полно – 2 балла• Частично – 1 балл• не раскрыта – 0 баллов <i>Логика изложения информации</i> <ul style="list-style-type: none">• логичное изложение материала – 2 балла• нарушение логики – 1 балл• отсутствие логики – 0 баллов <i>Использование ресурсов</i> <ul style="list-style-type: none">• использовано не менее 5 ресурсов – 2 балла• использовано 1-4 ресурса – 1 балл• не использован ни один ресурс – 0 баллов

<p><i>Продвинутый</i></p>	<p><i>Перевод текста с использованием электронных словарей</i></p>	<ul style="list-style-type: none"> • <i>Текст успешно переведен, слова верно подобраны по значению, выбрана правильная грамматическая форма, соблюден порядок слов – 3 балла</i> • <i>Текст содержит незначительные ошибки лексического или грамматического плана – 2 балла</i> • <i>Текст содержит большое количество лексических, грамматических, орфографических ошибок – 0 баллов</i>
<p>уметь решать лингвистические задачи с помощью электронных словарей и ресурсов <i>Базовый</i></p> <p><i>Продвинутый</i></p>	<p><i>Перевод заданных лексем с помощью электронных словарей</i></p> <p><i>Выявление с помощью программы типа «конкордансер» контекстов заданного слова, зарегистрированных в выбранном корпусе текстов соответствующего языка</i></p>	<ul style="list-style-type: none"> • <i>Лексема переведена – 1 балл</i> • <i>Лексема не переведена – 0 баллов</i> • <i>Контексты выявлены – 1 балл</i> • <i>Контексты не выявлены – 0 баллов</i>
<p>знать основные типы электронных лингвистических ресурсов <i>Базовый</i></p>	<p><i>Тест закрытого типа (множественного выбора, альтернативного выбора)</i></p>	<ul style="list-style-type: none"> • <i>Правильно выбран вариант ответа – 1 балл</i> • <i>Неправильно выбран вариант ответа – 0 баллов</i>

<i>Продвинутый</i>	<i>Развернутое устное или письменное высказывание</i>	<ul style="list-style-type: none"> • Тема раскрыта с опорой на соответствующие понятия и теоретические положения – 2 балла • Аргументация неполная, смысл ряда ключевых понятий не объяснен – 1 балл • Терминологический аппарат непосредственно не связан с раскрываемой темой – 0 баллов
--------------------	-------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

3. Типовые контрольные задания для проверки уровня сформированности компетенции 2 (ОПК-14 – владение основами современной информационной и библиографической культуры)

Этап формирования компетенции, в котором участвует дисциплина	Типовые контрольные задания для оценки знаний, умений, навыков (2-3 примера)	Показатели и критерии оценивания компетенции, шкала оценивания
<p>владеть навыками информационного поиска в сети Интернет; нормами оформления библиографических данных</p> <p><i>Базовый</i></p>	<p><i>Оформление списка литературы в соответствии со стандартом</i></p>	<ul style="list-style-type: none"> • Список литературы оформлен в соответствии с образцом – 1 балл Список литературы оформлен с нарушениями – 0 баллов
<p><i>Продвинутый</i></p>	<p><i>Написание реферата</i></p>	<p>Полнота раскрытия темы</p> <ul style="list-style-type: none"> • Полно – 2 балла • Частично – 1 балл

	<p><i>Поиск информации в поисковых системах по заданным параметрам</i></p>	<ul style="list-style-type: none"> • не раскрыта – 0 баллов <p>Логика изложения информации</p> <ul style="list-style-type: none"> • логичное изложение материала – 2 балла • нарушение логики – 1 балл • отсутствие логики – 0 баллов <p>Использование ресурсов</p> <ul style="list-style-type: none"> • использовано не менее 5 ресурсов – 2 балла • использовано 1-4 ресурса – 1 балл <p>не использован ни один ресурс – 0 баллов</p> <ul style="list-style-type: none"> • Искомая информация найдена – 1 балл • Информация не найдена – 0 баллов
<p><i>Продвинутый</i> уметь редактировать и форматировать текст в компьютерном текстовом редакторе в соответствии с поставленной задачей; формировать поисковый запрос</p> <p><i>Базовый</i></p>	<p><i>Оформление титульного листа реферата по дисциплине, оформленного по заданному шаблону</i></p>	<ul style="list-style-type: none"> • Титульный лист оформлен в соответствии с шаблоном (соблюдены параметры: расположение на странице, размер шрифта, выделение курсивом, разрядкой) – 1 балл • Оформление титульного листа не соответствует шаблону – 0 баллов
<p><i>Продвинутый</i></p>	<p><i>Редактирование текста – выбор шрифта, кегля, выделения текста, выравни-</i></p>	<ul style="list-style-type: none"> • Текст отредактирован в соответствии с заданными параметрами – 5 баллов

	<p>вания, междустрочного интервала, расстановка переносов и т.д.</p>	<p>(по баллу на параметр)</p> <ul style="list-style-type: none"> • Текст отредактирован не по всем заданным параметрам – вычитается балл за каждый несоблюденный параметр • Текст не отредактирован – 0 баллов
<p><i>Продвинутый</i> знать основные алгоритмы информационного поиска и базовые операторы языковых запросов; требования к оформлению библиографической информации.</p>		
<p><i>Базовый</i></p>	<p><i>Тест закрытого типа (множественного выбора, альтернативного выбора)</i></p>	<ul style="list-style-type: none"> • Правильно выбран вариант ответа – 1 балл • Неправильно выбран вариант ответа – 0 баллов
<p><i>Продвинутый</i></p>	<p><i>Развернутое устное или письменное высказывание</i></p>	<ul style="list-style-type: none"> • Тема раскрыта с опорой на соответствующие понятия и теоретические положения – 2 балла • Аргументация неполная, смысл ряда ключевых понятий не объяснен – 1 балл • Терминологический аппарат непосредственно не связан с раскрываемой темой – 0 баллов

4. Типовые контрольные задания для проверки уровня сформированности компетенции 3 (ОПК-20 – способность решать стандартные задачи професси-

ональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учётом основных требований информационной безопасности)

<p>Этап формирования компетенции, в котором участвует дисциплина</p>	<p>Типовые контрольные задания для оценки знаний, умений, навыков (2-3 примера)</p>	<p>Показатели и критерии оценивания компетенции, шкала оценивания</p>
<p>владеть навыками решения стандартных профессиональных задач с применением информационно-лингвистических технологий и электронных ресурсов <i>Базовый</i></p>	<p><i>Работа с корпусом английского / немецкого / французского / русского языка: различная формулировка запросов</i></p> <p><i>Поиск записей по ключевому слову в электронном корпусе личных дневников XX века</i></p>	<ul style="list-style-type: none"> • <i>Запрос выполнен в соответствии с заданием – 1 балл</i> • <i>Запрос выполнен некорректно – 0 баллов</i> • <i>Поиск выполнен – 1 балл</i> • <i>Поиск не выполнен – 0 баллов</i>
<p>уметь формализовать языковой материал для решения поставленной задачи; проводить отбор и анализ языковых фактов с целью применения информационных технологий <i>Базовый</i></p>	<p><i>Выбор подкорпуса национального корпуса текстов в зависимости от цели исследования</i></p> <p><i>Работа с национальным корпусом текстов: сортировка примеров употребления заданного слова по дате создания</i></p>	<ul style="list-style-type: none"> • <i>Выбор подкорпуса соответствует цели исследования – 1 балл</i> • <i>Выбор подкорпуса не отвечает цели исследования – 0 баллов</i> • <i>Сортировка проведена – 1 балл</i> • <i>Сортировка не проведена – 0 баллов</i>
<p>знать основные области применения информационных технологий в лингвистике; терминологический аппарат, связанный с использованием информационных технологий в лингвистике; основные составляющие ин-</p>		

<p>формационно-коммуникационной компетенции лингвиста. Базовый</p>	<p><i>Тест закрытого типа (множественного выбора, альтернативного выбора)</i></p>	<ul style="list-style-type: none"> • <i>Правильно выбран вариант ответа – 1 балл</i> • <i>Неправильно выбран вариант ответа – 0 баллов</i>
------------------------------------------------------------------------	-----------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------

V. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (или модуля)

а) Основная литература:

1. Исакова А. И., Исаков М. Н. Информационные технологии : учеб. пособие. – Томск: Эль Контент, 2012. – 174 с. Режим доступа в Университетской библиотеке ONLINE:

http://biblioclub.ru/index.php?page=book_view_red&book_id=208647

2. Копотев М. Введение в корпусную лингвистику : учеб. пособие для студентов филолог. и лингвист. специальностей университетов. – Прага: Animedia Company, 2014. – 195 с. Режим доступа в Университетской библиотеке ONLINE: <http://biblioclub.ru/index.php?page=book&id=375463>

б) Дополнительная литература:

1. Фролов А., Фролов Г. Синтез и распознавание речи. Современные решения [электронный учебник]. Режим доступа: <http://www.frolov-lib.ru/books/hi1/index.html>

2. Информатика и лингвистика: учебное пособие / Волосатова Т.М., Чичварин Н.В. - М.: НИЦ ИНФРА-М, 2016. - 196 с.: 60x90 1/16. - (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977-0. - Режим доступа: <http://znanium.com/go.php?id=508097>

3. Базовые и прикладные информационные технологии: Учебник / Гвоздева В. А. - М.: ИД ФОРУМ, НИЦ ИНФРА-М, 2015. - 384 с.: 60x90 1/16. - (Высшее образование) (Переплёт 7БЦ) ISBN 978-5-8199-0572-2. – Режим доступа: <http://znanium.com/catalog.php?bookinfo=504788>

4. Шишов О.В. Современные технологии и технические средства информатизации: Учебник. — М.: ИНФРА-М, 2017. — 462 с. — (Высшее образование: Бакалавриат). — Режим доступа: <http://znanium.com/catalog.php?bookinfo=653093>

VI. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (или модуля)

1. Ассоциация компьютерной лингвистики URL: <http://aclweb.org/>
2. Библиотека художественной литературы URL: <http://www.lib.com.ua>
3. Британский национальный корпус URL: <http://info.ox.ac.uk/bnk>
4. МультиЛекс URL: <http://www.multilex.ru>
5. Национальный корпус русского языка URL: <http://www.ruscorpora.ru>
6. Русская виртуальная библиотека URL: <http://www.rvb.ru>
7. Словари на Академике URL: <http://dic.academic.ru>
8. Электронная библиотека URL: <http://feb-web.ru>
9. Электронная библиотека URL: e-library.ru
10. Яндекс. Словари URL: <http://slovari.yandex.ru>
11. Bank of English URL: http://titania.cobuild.collins.co.uk/boe_info.html/
12. Oxford Text Archive URL: <http://sable.ox.ac.uk/ota/>
13. Wikipedia. The Free Encyclopedia URL: <http://www.wikipedia.org>
14. ЭБС «Университетская библиотека онлайн» URL: biblioclub.ru

VII. Методические указания для обучающихся по освоению дисциплины (или модуля)

ТЕМАТИЧЕСКОЕ СОДЕРЖАНИЕ КУРСА

ОСНОВНЫЕ ПОНЯТИЯ КУРСА «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЛИНГВИСТИКЕ»

Общее понятие «информационные технологии». «Электронный ресурс» как общее понятие. Виды электронных ресурсов и принципы их классификации. Фундаментальная лингвистика, прикладная лингвистика и компьютерная лингвистика. Основные направления прикладной/компьютерной лингвистики.

Различные трактовки понятия «гипертекст». Гипертекст и особенности переработки информации человеком. Электронный гипертекст как особая форма текста, принципиально отличающаяся от бумажных текстов. Роль гипертекста в познании и в образовании (обучении). Виды ссылок; гиперссылки и их особенности и преимущества.

СОЗДАНИЕ И ИСПОЛЬЗОВАНИЕ БАЗ ДАННЫХ

«База данных» как совокупность определенным образом упорядоченных сведений о некоторых объектах. Типы данных. Способы организации базы данных. Формирование базы данных для научно-исследовательской работы. Формирование базы данных на основе многоязычных электронных словарей. Возможности количественного и качественного анализа отобранного корпуса данных. Представление результатов статистической обработки и продуктов качественного анализа материалов с помощью рисунков, диаграмм и т.д. Формиро-

вание базы данных на основе корпусов различных видов: постановка задачи; критерии отбора текстов; параметры анализа отобранных данных.

ОСНОВЫ ИНФОРМАЦИОННОГО ПОИСКА И БИБЛИОГРАФИЧЕСКОЙ КУЛЬТУРЫ.

Источники информации в сети Интернет. Основные понятия информационного поиска (релевантность, полнота, точность, информационный шум). Типы информационно-поисковых систем. Язык поисковых запросов и алгоритмы поиска. Стандарты библиографического описания источников информации.

АВТОМАТИЧЕСКИЙ АНАЛИЗ И СИНТЕЗ ЗВУЧАЩЕЙ РЕЧИ; АВТОМАТИЧЕСКИЙ АНАЛИЗ И СИНТЕЗ ТЕКСТА.

Понятия автоматического анализа и синтеза речи и текста: проблемы и сферы применения. Этапы автоматического анализа речи. Аналоговый и цифровой звуковой сигнал. Пословный и фонемный анализ речи. Методы автоматического синтеза речи. Графематический, морфологический, синтаксический и семантический анализ текста. Понятия токенизации, лемматизации, парсинга. Формальная грамматика. Автоматический синтез текста.

МАШИННЫЙ ПЕРЕВОД И АВТОМАТИЗИРОВАННЫЙ ПЕРЕВОД.

Машинный перевод vs автоматизированный перевод. Краткая история машинного перевода. Машинный перевод, основанный на правилах: принципы работы, преимущества и недостатки. Статистический машинный перевод: принципы работы, преимущества и недостатки. Технология памяти переводов (Translation Memory): принципы работы, наиболее распространённые программы.

ЭЛЕКТРОННАЯ ЛЕКСИКОГРАФИЯ.

Основные типы словарей и параметры их классификации. Недостатки традиционных бумажных словарей и преимущества электронного словаря как особого лексикографического объекта. Особенности электронного словаря, существующего в среде многомерного гипертекста. Типы современных электронных словарей. Электронные библиотеки. Электронные энциклопедии. Конкордансы. Ассоциативные тезаурусы и возможности их использования для разных целей. Возможности использования одноязычных и многоязычных электронных словарей для формирования корпуса данных при проведении исследований в различных целях.

КОРПУСНАЯ ЛИНГВИСТИКА.

Корпусная лингвистика как область прикладной лингвистики. Понятие корпуса текстов. Требования к корпусу текстов с точки зрения пользователя. Основные виды корпусов и их специфика. Типы корпусной разметки.

Типы корпусов текстов и основания для их классификации. Возможности использования корпусов текстов для проведения лингвистических исследова-

ний.

ГЛОССАРИЙ ОСНОВНЫХ ТЕРМИНОВ К КУРСУ «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЛИНГВИСТИКЕ»

Автоматизированный перевод (CAT – computer-assisted translation): перевод текстов человеком с использованием компьютерных технологий.

Автоматический анализ звучащей речи: преобразование звучащей речи в печатный текст, над которым можно производить дальнейшие операции.

Автоматический синтез звучащей речи: процесс преобразования печатного текста, существующего в цифровой форме, в звучащий текст на естественном человеческом языке.

Автоматический анализ текста: последовательное преобразование текста на естественном человеческом языке, введенного в компьютер, в его лексемно-морфологические, синтаксические и семантические представления, понятные компьютеру.

Автоматический синтез текста: процесс преобразования лексемно-морфологических, синтаксических и семантических представлений в текст на естественном языке.

База данных: поименованная совокупность структурированных данных некоторой предметной области, которая хранится на некотором материальном носителе.

Бот, интернет-бот: программа, выполняющая автоматически и/или по заданному расписанию какие-либо действия через интерфейсы, предназначенные для людей (напр., робот-автоответчик, чат-бот, бот в компьютерных играх).

Гипертекст: набор семантически связанных блоков, которые существуют в компьютерной среде и объединяются в сеть посредством ассоциативно или иерархически организованных гиперссылок.

Дерево зависимостей: формальная модель представления синтаксической структуры предложения, применяющаяся в автоматическом синтаксическом анализе и в системах машинного перевода, основанных на правилах. Все связи в предложении рассматриваются как подчинительные, а исходным пунктом анализа выступает сказуемое, находящееся в вершине графа.

Дерево непосредственно составляющих: формальная модель представления синтаксической структуры предложения, применяющаяся в автоматическом синтаксическом анализе и в системах машинного перевода, основанных на правилах. Данный тип членения соответствует разбиению структуры на пары элементов: основной состав предложения разбивается на группу подлежащего и группу сказуемого.

Информационно-поисковая система: упорядоченная совокупность документов и информационных технологий, предназначенных для хранения и поиска информации в виде целых текстов или отдельных упоминаемых в них фактов.

Информационно-поисковый язык: формальный язык, предназначенный для описания содержания документов, хранящихся в информационно-поисковой системе, и запроса пользователя.

Информационные технологии: компьютерные инструменты получения, хранения, передачи, распространения и преобразования информации, а также соответствующие законы и методы.

Информационные технологии в лингвистике: компьютерные инструменты получения, хранения, передачи, распространения и преобразования информации о языке и законах его функционирования, а также соответствующие законы и методы.

Информационный шум: множество документов, выдаваемых в процессе информационного поиска, формально соответствующих запросу (релевантных), но не являющихся релевантными по смыслу.

Информация: сведения об окружающем мире, передаваемые человеком, живыми организмами или техническими системами для регулирования своего поведения в окружающей среде.

Искусственные языки: знаковые системы, искусственно создаваемые в тех областях, где применение естественных языков менее эффективно или не возможно.

Искусственный интеллект: направление в информатике, которое занимается созданием интеллектуальных когнитивных технологий и машин, способных понимать, моделировать и анализировать тексты, хранить и перерабатывать естественно-языковую информацию, принимать интеллектуальные решения.

Кодирование: процесс представления информации в виде последовательности условных обозначений; сопоставление объектов и отношений между ними с символами или словами какого-либо языка.

Компьютерная (электронная) лексикография: раздел прикладной лингвистики, нацеленный на создание электронных (автоматических) словарей, лингвистических баз данных и разработку программ поддержки лексикографических работ.

Компьютерная лингвистика (англ. computational linguistics): область использования компьютерных инструментов — программ, технологий организации и обработки данных — для моделирования функционирования языка в тех или иных условиях, а также сферу применения компьютерных моделей языка в лингвистике и смежных с ней дисциплинах.

Конкорданс: список всех контекстов употребления искомой единицы в корпусе.

Конкордансер: электронная программа, помогающая получить список употреблений слова или словосочетания в выбранном корпусе для установления возможных контекстов.

Корпус лингвистический: совокупность специально отобранных текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска.

Корпусная лингвистика: раздел прикладной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов при помощи компьютеров.

Лемма: единица словаря, начальная форма слова, к которой можно привести все отдельные словоупотребления и грамматические формы одной лексемы.

Машинная основа: последовательность букв от начала словоформы, общая для всех словоформ, входящих в формообразовательную парадигму данного слова, например (рус.) блок#, включ# или (англ.) buil#, earl#.

Машинное окончание: элемент, описывающий формоизменение конкретной лексемы; машинные окончания представляются в виде парадигм.

Машинный перевод: процесс перевода текстов с одного естественного языка на другой специальной компьютерной программой.

Модель: материальный или идеальный образ некоторой совокупности предметов или явлений, заменяющий реальные предметы и явления и включающий их наиболее существенные признаки. В лингвистике модели имитируют строение и функционирование языка, производство и восприятие речи и текста.

Модель данных: логическая структура представления данных, хранимых в базе, и операций манипулирования ими (например, иерархическая, сетевая, реляционная).

n-грамма: последовательность из n элементов (биграмма – последовательность из двух элементов; триграмма – последовательность из трёх элементов; 4-грамма – последовательность из пяти элементов и т.д.)

Омонимия: совпадение различных языковых единиц, значения которых не связаны друг с другом.

Морфологическая омонимия – совпадение грамматических форм одного и того же слова.

Лексико-морфологическая омонимия – совпадение словоформ двух разных лексем.

Синтаксическая омонимия – неоднозначность синтаксической структуры, имеющей несколько интерпретаций.

Основа: ядерная часть слова без словоизменятельных морфем.

Парсер: специальная компьютерная программа для автоматического анализа слов, морфологического или синтаксического.

Поисковая система: (автоматизированная) компьютерная система, позволяющая находить веб-ресурсы, содержание которых соответствует запросу пользователя.

Поле объекта OLE (Object Linking and Embedding): тип поля в таблице базы данных (например, Microsoft Access), позволяющий установить связь с объектами другого приложения или внедрить объект в базу данных. Объектами могут быть простые и форматированные тексты, рисунки, диаграммы, аудио- и видеофайлы и т.д.

Полнота поиска: соотношение между количеством выданных релевантных (по отношению к конкретному запросу) документов к общему числу релевантных документов, имеющихся в информационной система. Обычно выражается в процентах.

Прикладная лингвистика: научная дисциплина, в которой изучаются и разрабатываются способы оптимизации функционирования языка в нелингвистических научных дисциплинах и в различных сферах практической деятельности человека; отвечает на вопрос «Как лучше использовать язык?».

Разметка (тэггинг): приписывание текстам и их компонентам специальных меток (тэгов).

Релевантность: соответствие выданного документа поисковому запросу.

Система переводческой памяти (Translation Memory, TM): программа, сохраняющая переводы, сделанные ранее, и предлагающая человеку уже готовый перевод фразы или фрагмента текста, если он уже был однажды переведён.

Система управления базой данных (СУБД): совокупность программных средств, позволяющих осуществить ведение базы данных (создание, обновление, удаление элементов и т.д.) и поиск информации в базе данных.

Снятие омонимии (=разрешение многозначности): выбор правильной интерпретации словоформы, допускающей несколько вариантов толкований.

Тег (тэг): 1) специальный символ разметки языка HTML, позволяющий визуально оформить содержимое веб-сайтов (напр., начало нового абзаца, шрифтовое выделение текста, оформление гиперссылки и т.д.);

2) метка статьи, записи в блоге или любого другого элемента, которая служит для группировки элементов, помеченных одним и тем же тегом;

3) особый код в лингвистической разметке корпуса текстов, соответствующий определённому признаку или набору признаков слова.

Тест Тьюринга: эмпирический тест, предложенный Аланом Тьюрингом (1950), суть которого (в упрощённом виде) сводится к тому, что несколько участников беседуют с некоторым участником X, и если большинство участников не сможет установить, с кем они беседуют – с человеком или машиной, проблема построения искусственного интеллекта решена.

Типовая парадигма лексемы: все возможные наборы машинных окончаний лексем.

Токен: отдельное словоупотребление, слово от пробела до пробела.

Точность поиска: соотношение между количеством выданных релевантных документов к общему количеству документов в выдаче (обычно выражается в процентах).

Форманта: термин фонетики, обозначающий акустическую характеристику звуков речи (прежде всего гласных), связанную с уровнем частоты голосового тона и образующую тембр звука. Представляет достаточно отчётливо выделяющуюся область усиленных частот в спектре звука.

Формат KWIC (key word in context): формат представления конкорданса в корпусе, при котором примеры употреблений выровнены по ключевому слову, расположенному в центре строки.

Электронный (автоматический) словарь: собрание слов и их комментариев в специальном машинном формате, предназначенное для использования человеком или являющееся составной частью более сложных компьютерных программ (например, систем машинного перевода).

БАНК КОНТРОЛЬНЫХ ВОПРОСОВ ДЛЯ ПОДГОТОВКИ К ПРАКТИЧЕСКИМ (ЛАБОРАТОРНЫМ) ЗАНЯТИЯМ И ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В чём состоят различия между компьютеризацией общества и информатизацией общества?

Что понимается под информационными ресурсами?

Какие виды материальных носителей информации вам известны?

Что может входить в состав информационных ресурсов?

Что понимается под информационными технологиями?

Что понимается под информационными технологиями в лингвистике?

Каковы цели применения информационных технологий в лингвистике?

В чём состоят преимущества применения новых информационных технологий в лингвистике?

Какие определения понятия «база данных» Вам известны?

Что понимается под структурированием базы данных?

Какие типы данных могут использоваться в базах данных?

Какие способы организации базы данных Вам известны?

В чём состоят особенности иерархической модели организации базы данных?

В чём состоят особенности сетевой модели организации базы данных?

В чём состоят особенности реляционной модели организации базы данных?

Каковы типы связей (отношения) между составляющими базы данных в различных моделях организации базы данных?

Что понимается под системой управления базой данных?

Какие виды систем управления базой данных Вам известны?

Какие формы существования информационных ресурсов Вам известны?

Что такое «гипертекст»?

Каковы основные признаки гипертекста?

Какие тексты имеют некоторые признаки гипертекста?

В чём состоят принципиальные различия между бумажным текстом и электронным гипертекстом?

Каковы основные особенности гипертекста?

Какие типы структуры гипертекста Вам известны?

Какие виды гипертекста Вам известны?

Чем различаются электронные словари и компьютерные словари?

В чём состоят недостатки печатных словарей?

В чём состоят преимущества электронных словарей?

Какая связь существует между электронным словарём и идеей гипертекста?

На каких основаниях может строиться типология словарей?

Чем различаются частотно-алфавитный словарь и словоуказатель?

Что такое конкорданс?

Что такое конкордансер?

Что такое тезаурус?

Какие виды тезаурусов Вам известны?

Что из себя представляет энциклопедия?

Какие энциклопедии являются самыми известными?

Какие Вы знаете мультимедийные электронные энциклопедии?

Какие новые дополнительные свойства появляются у электронных словарей и энциклопедий?

Какие крупные каталоги электронных словарей Вам известны?

Какие терминологические значения слова «корпус» Вам известны?

Что может пониматься под корпусом текстов?

Чем занимается корпусная лингвистика?

Что Вам известно из истории разработки больших корпусов?

Что понимается под «большим корпусом»?

Как называется пополняемый корпус текстов?

Какие национальные корпуса текстов Вам известны?

Что понимается под параллельным корпусом?

В чём заключается разница между выровненными и невыровненными корпусами?

По каким основаниям могут классифицироваться корпуса текстов?

Что такое корпусная разметка?

Какие виды разметки текстов Вам известны?

Что входит во внешелингвистическую разметку текстов?

Что входит в лингвистическую разметку текстов?

Что входит в морфологическую разметку текстов?

Что входит в синтаксическую разметку текстов?

Что входит в семантическую разметку текстов?

Окно документа, работа с несколькими окнами. Управление панелями инструментов.

Основные элементы текста (символ, слово, приложение, абзац). Операции с текстом: ввод, работа с фрагментами (выделение, копирование, перенос, удаление).

Методы форматирования элементов текста (тип, размер, вид шрифта), размещение на странице в одну, две и более колонки.

Как используется стиль символов для форматирования шрифта? Как и с какой целью применяется стиль абзацев в MS Word?

Форматирование символов и абзацев в текстовом процессоре MS Word. Операции копирования, переноса, удаления фрагментов и роль буфера обмена.

Поиск и замена фрагментов текста, использование автотекста, автопереноса.

Создание и форматирование таблиц в процессоре MS WORD

Работа с рисунками в текстовом процессоре MS WORD.

Что такое компьютерные сети. Как организована всемирная сеть Интернет? Web-сервер и Web-браузер.

Назначение и основные возможности табличного процессора MS Excel.

Структура электронной таблицы, ее основные понятия.

Формула. Мастер функций.

Содержание адреса Web-страниц и принципы адресации URL. Какие поисковые технологии использует Интернет?

Какие способы передачи электронных сообщений использует e-mail? Что может входить в состав электронного сообщения? Как переслать файлы? Этика письма.

Структура адреса Web-страниц и принципы адресации URL.

БАНК ТИПОВЫХ ЗАДАНИЙ ДЛЯ ПРАКТИЧЕСКИХ (ЛАБОРАТОРНЫХ) ЗАНЯТИЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Наберите в текстовом редакторе Word текст и отформатируйте его по заданным преподавателем параметрам.

Представьте титульный лист реферата по дисциплине, оформленный по заданному шаблону.

Представьте список литературы к реферату по дисциплине, оформленный по заданному шаблону.

Проведите анализ лексической единицы с использованием словарей различных типов.

Используя поисковые системы Интернета, установите заглавие и автора предложенного текста.

Используя поисковые системы Интернета, найдите точные цитаты на исходном языке для предложенного текстового фрагмента.

Установите, какие возможности для получения статистических данных обеспечивает

разметка текстов выбранного Вами подкорпуса.

С помощью программы типа «конкордансер» выявите контексты заданного слова, зарегистрированные в выбранном корпусе текстов соответствующего языка.

Постройте подкорпус по определённой тематике.

Постройте диаграмму по результатам поиска.

Определите процентное соотношение текстов в заданном корпусе.

Проследите динамику использования заданных лексических единиц в подкорпусе / подкорпусах.

Создайте гиперссылки в предложенном документе.

Создайте изолированный гипертекст.

Создайте сетевой гипертекст.

ТЕМАТИКА РЕФЕРАТОВ

Основные направления прикладной лингвистики.

Цели и задачи использования информационных технологий в лингвистике.

Принципы классификации электронных ресурсов.

Гипертекст: признаки, виды, структура.

Виды электронных словарей.

Особенности электронных словарей.

Особенности электронных библиотек.

Основные электронные библиотеки в РФ.

Основные электронные библиотеки в мире.

Мультимедийные энциклопедии.

Специфика корпусной лингвистики.

Понятие «Большой корпус национального языка».

Основные корпуса национальных языков.

Структурная организация корпусов.

Типы корпусной разметки.

Возможности использования корпусов национальных языков для решения учебно-методических задач.

Возможности использования корпусов национальных языков для решения научно-исследовательских задач.

Возможности использования корпусов национальных языков в переводе.

Сравнительный анализ составления поисковых запросов в популярных русскоязычных поисковых системах.

Обзор сетевых ресурсов по прикладной лингвистике / корпусной лингвистике / автоматизированному переводу.

ЭКЗАМЕНАЦИОННЫЕ ВОПРОСЫ ПО ДИСЦИПЛИНЕ «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЛИНГВИСТИКЕ»

1. Информационные ресурсы и технологии в лингвистике: цели, виды, преимущества.

2. Фундаментальная и прикладная лингвистика. Основные направления прикладной/компьютерной лингвистики.

3. Автоматический анализ и синтез звучащей речи: этапы, методы, сферы применения.

4. Автоматический анализ и синтез текста: этапы, проблемы, сферы примене-

ния.

5. Базы данных: способы организации, типы данных.

6. Корпусная лингвистика: типы корпусов, возможности использования.

7. Корпусная лингвистика: виды корпусной разметки, возможности использования.

8. Электронная лексикография: недостатки и преимущества электронных словарей.

9. Электронная лексикография: виды словарей, основания для классификации.

10. Машинный перевод, основанный на правилах: принципы работы, преимущества и недостатки.

11. Статистический машинный перевод: принципы работы, преимущества и недостатки.

12. Машинный перевод vs автоматизированный перевод. Технология памяти переводов (Translation Memory): принципы работы, названия программ.

ТРЕБОВАНИЯ К РЕЙТИНГ-КОНТРОЛЮ

Форма промежуточной аттестации – **экзамен**.

Распределение баллов (макс. 100):

Работа в семестре – до 60 баллов, включая

- посещаемость и работа на занятиях – до 40 баллов;
- модульный контроль – до 20 баллов.

Экзамен – до 40 баллов, включая

- итоговый тест – до 20 баллов;
- устный развёрнутый ответ на теоретический вопрос – до 10 баллов;
- практическое задание на проверку сформированности компетенций – до 10 баллов.

VIII. Перечень педагогических и информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (или модулю), включая перечень программного обеспечения и информационных справочных систем (по необходимости)

При организации и проведении занятий по дисциплине используются следующие образовательные технологии:

- различные типы лекций (традиционная, проблемная, дискуссия)

- информационно-поисковая технология
- постановка и решение проблемных задач
- практические задания.

IX. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (или модулю)

Учебная аудитория № 206 (170100, Тверская область, г. Тверь, ул. Желябова, д.33)	Экран, проектор, компьютер	Google Chrome – бесплатно Kaspersky Endpoint Security 10 для Windows – Акт на передачу прав №2129 от 25 октября 2016 г. MS Office 365 pro plus - Акт приема-передачи № 369 от 21 июля 2017 Microsoft Windows 10 Enterprise - Акт приема-передачи № 369 от 21 июля 2017
----------------------------------------------------------------------------------	----------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

X. Сведения об обновлении рабочей программы дисциплины (или модуля)

№п.п.	Обновленный раздел рабочей программы дисциплины (или модуля)	Описание внесенных изменений	Дата и протокол заседания кафедры, утвердившего изменения
1.	РПД	Обновление макета	13.06.2017, протокол № 10
2.	Раздел 4	Актуализация ФОСов	13.06.2017, протокол № 10
3.	Раздел 5	Актуализация литературы	13.06.2017, протокол № 10
4.	Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине	Актуализация списка МТО и ПО	Совет факультета ИЯ и МК от 5.09.2017