

УТВЕРЖДАЮ
 Декан факультета

 (подпись) Шматко А.Д.
 ФИО
 «___» _____ 20__

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ТЕХНОЛОГИИ КОРПУСНОЙ ЛИНГВИСТИКИ

Направление/специальность подготовки	45.05.01 Перевод и переводоведение
Специализация/профиль/программа подготовки	Специальный перевод
Уровень высшего образования	Специалитет
Форма обучения	Очная
Факультет	Р Международного промышленного менеджмента и коммуникации
Выпускающая кафедра	Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА
Кафедра-разработчик рабочей программы	Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА

КУРС	СЕМЕСТР	ОБЩАЯ ТРУДОЁМКОСТЬ (ЗАЧЕТНЫХ ЕДИНИЦ)	ЧАСЫ (по наличию видов занятий)									ВИД ПРОМЕЖУТОЧНОГО КОНТРОЛЯ
			ОБЩАЯ ТРУДОЁМКОСТЬ	АУДИТОРНЫЕ ЗАНЯТИЯ				САМОСТОЯТЕЛЬНАЯ РАБОТА				
				ВСЕГО	ЛЕКЦИИ	ЛАБОРАТОРНЫЙ ПРАКТИКУМ	ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	ВСЕГО	КУРСОВОЙ ПРОЕКТ	КУРСОВАЯ РАБОТА	ДРУГИЕ ВИДЫ САМОСТ. РАБОТЫ	
5	9	3	108	34	17	0	17	74	0	0	74	ЭКЗ.

ЛИСТ СОГЛАСОВАНИЯ

РАБОЧАЯ ПРОГРАММА СОСТАВЛЕНА В СООТВЕТСТВИИ С ТРЕБОВАНИЯМИ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО ОБРАЗОВАТЕЛЬНОГО СТАНДАРТА ВЫСШЕГО ОБРАЗОВАНИЯ (ФГОС ВО)

45.05.01 Перевод и переводоведение

год набора группы: 2025

Программу составил:

Кафедра Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА
Мамаев Иван Дмитриевич, к.ф.н., старший преподаватель

Программа рассмотрена
на заседании кафедры-разработчика
рабочей программы **Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА**

Заведующий кафедрой Невзорова Г.Д., к.ф.н., доц.

Программа рассмотрена
на заседании выпускающей кафедры

Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА

Заведующий кафедрой Невзорова Г.Д., к.ф.н., доц.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ТЕХНОЛОГИИ КОРПУСНОЙ ЛИНГВИСТИКИ

Разделы рабочей программы

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО
3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ
4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ
5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Приложения к рабочей программе дисциплины

- Приложение 1. Аннотация рабочей программы
- Приложение 2. Технологии и формы обучения
- Приложение 3. Фонды оценочных средств

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-4 — Способен работать с электронными словарями, различными источниками информации, осуществлять поиск, хранение, обработку и анализ информации, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий

Формированию компетенций служит достижение следующих результатов образования:

ОПК-4

знания:

понятийная база и терминологический минимум корпусной лингвистики;
основные типы корпусов, требования, предъявляемые к корпусам;
сущность процедуры разметки и основные стандарты корпусной разметки;

умения:

работа с программными средствами и информационными ресурсами корпусной лингвистики;

навыки:

работа с корпусными базами данных в целях лингвистического анализа;
осуществление поиска и исследований на базе корпусов.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина **ТЕХНОЛОГИИ КОРПУСНОЙ ЛИНГВИСТИКИ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению *45.05.01 Перевод и переводоведение*.

Содержание дисциплины является логическим продолжением дисциплин: **ОБЩАЯ МОРФОЛОГИЯ, ЛЕКСИКОЛОГИЯ**.

Содержание дисциплины является основой для освоения дисциплин: **ДИСКУРС**.

Предварительные компетенции, сформированные у обучающегося до начала изучения дисциплины:

- ОПК-1 — Способен применять знания иностранных языков и знания о закономерностях функционирования языков перевода, а также использовать систему лингвистических знаний при осуществлении профессиональной деятельности

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч.

3.1. Содержание (дидактика) дисциплины

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме			Самостоятельная работа студентов	Формируемая компетенция, %
				ВСЕГО	Лекции	Практические занятия		ОПК-4
5	9	Раздел 1. Введение в корпусную лингвистику. Классификация корпусов. Языковой материал в лингвистическом исследовании. Общее представление о корпусах и корпусной лингвистике. Основные понятия. Основные требования, предъявляемые к корпусам. Компьютерная поддержка корпусов, программы для работы с корпусами. Принципы классификации корпусов. Типы корпусов. Наиболее известные корпусы.	30	10	5	5	20	25
5	9	Раздел 2. История развития корпусной лингвистики. Национальные корпусы текстов. Зарождение и развитие корпусной лингвистики. Опыт создания корпусов в США и в Европе. Национальные корпусы России, Западной Европы и США. Характеристика корпуса на примере Национального корпуса русского языка.	30	10	5	5	20	25
5	9	Раздел 3. Стандарты и типы разметки в корпусах. Аннотирование корпусов. Принципы разметки. Виды корпусной разметки. Морфологическая разметка корпуса. Синтаксическая разметка корпуса. Семантическая разметка корпуса.	28	6	3	3	22	25
5	9	Раздел 4. Сферы практического использования корпусов. Корпусные методы исследований. Сферы применения корпусов. Особенности применения корпусов текстов. Применение корпусов в гуманитарных исследованиях. Лингвистические исследования, базирующиеся на корпусах. Проблемы и перспективы развития корпусной лингвистики.	20	8	4	4	12	25
Всего за 9 семестр			108	34	17	17	74	100
Всего по дисциплине			108	34	17	17	74	100

3.2. Аудиторный практикум

№ п/п	Номер и наименование раздела дисциплины	Тема практического занятия	Объем, ауд. часов
1	Раздел 1. Введение в корпусную лингвистику. Классификация корпусов.	Языковой материал в лингвистическом исследовании. Общее представление о корпусах и корпусной лингвистике. Основные понятия. Основные требования, предъявляемые к корпусам. Компьютерная поддержка корпусов, программы для работы с корпусами.	2
2		Принципы классификации корпусов. Типы корпусов. Наиболее известные корпусы.	3
3	Раздел 2. История развития корпусной лингвистики.	Характеристика корпуса на примере Национального корпуса русского языка.	3
4	Национальные корпусы текстов.	Зарождение и развитие корпусной лингвистики. Опыт создания корпусов в США и в Европе. Национальные корпусы России, Западной Европы и США.	2
5	Раздел 3. Стандарты и типы разметки в корпусах.	Аннотирование корпусов. Принципы разметки. Виды корпусной разметки.	1
6		Морфологическая разметка корпуса. Синтаксическая разметка корпуса. Семантическая разметка корпуса.	2
7	Раздел 4. Сферы практического	Лингвистические исследования, базирующиеся на корпусах.	1
8	использования корпусов. Корпусные методы исследований.	Проблемы и перспективы развития корпусной лингвистики.	1
9		Сферы применения корпусов. Особенности применения корпусов текстов. Применение корпусов в гуманитарных исследованиях.	2
Всего за 9 семестр			17

3.3. Самостоятельная работа студента (СРС)

№ п/п	Номер и наименование раздела дисциплины	Содержание учебного задания	Объем, часов
1	Раздел 1. Введение в корпусную лингвистику. Классификация	Изучение темы 1 по конспектам лекций и рекомендованной литературе для подготовки к	20

	корпусов.	практическим занятиям. Подготовка докладов.	
2	Раздел 2. История развития корпусной лингвистики. Национальные корпуса текстов.	Изучение темы 2 по конспектам лекций и рекомендованной литературе для подготовки к практическим занятиям. Подготовка докладов.	20
3	Раздел 3. Стандарты и типы разметки в корпусах.	Изучение темы 3 по конспектам лекций и рекомендованной литературе для подготовки к практическим занятиям. Подготовка докладов.	22
4	Раздел 4. Сферы практического использования корпусов. Корпусные методы исследований.	Изучение темы 4 по конспектам лекций и рекомендованной литературе для подготовки к практическим занятиям. Подготовка докладов.	12
Всего за 9 семестр			74

4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

СЕМЕСТР	НЕДЕЛИ СЕМЕСТРА																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
9					Докл	ДР			Докл, Задан	ДР		Докл, Задан	Вопр. Зач		Докл	ДР	

Условные обозначения:

- ДР – диагностическая работа;
- Докл – доклад;
- Задан – задание;
- Вопр. Зач – вопросы к зачету.

Текущий контроль успеваемости студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- доклад;
- задание;
- вопросы к зачету.

Промежуточная аттестация проводится в формах:

- экзамен.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература по дисциплине:

1. А. А. Шунейко. . Корпусная лингвистика. Москва: Юрайт, 2022, эл. рес.
2. А. В. Гребенщикова. . Основы квантитативной лингвистики и новых информационных технологий. Москва: Флинта, 2018, эл. рес.
3. Е. В. Булдакова. . Компьютерная лексикография. СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2009, эл. рес.

5.2. Дополнительная литература по дисциплине:

1. А. Н. Баранов. . Введение в прикладную лингвистику. М.: Эдиториал УРСС, 2003, 2 экз.

5.3. Периодические издания:

не требуются.

5.4. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины, электронные библиотечные системы:

1. <https://e.lanbook.com/> — ЭБС Лань;
2. <https://urait.ru/> — Главная – Образовательная платформа Юрайт. Для вузов и ссузов.;
3. http://library.voenmeh.ru/jirbis2/index.php?option=com_irbis&view=irbis&Itemid=474 — Фундаментальная библиотека БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова.

Современные профессиональные базы данных:

1. <https://rusneb.ru> – Национальная электронная библиотека (НЭБ);
2. <https://cyberleninka.ru/> - Научная электронная библиотека «Киберленинка»;
<http://www.rfbr.ru/rffi/ru/library> - Полнотекстовая электронная библиотека Российского фонда фундаментальных исследований.

Информационные справочные системы:

1. Техэксперт – Информационный портал технического регулирования: Нормы, правила, стандарты РФ;
2. http://library.voenmeh.ru/jirbis2/index.php?option=com_irbis&view=irbis&Itemid=457 - БД ГОСТов собственной генерации БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова;
3. <http://www.consultant.ru/>- КонсультантПлюс- информационный портал правовой информации.

5.5. Программное обеспечение:

не требуется.

5.6. Информационные технологии:

взаимодействие с обучающимися посредством ЭИОС Moodle БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Лекционные занятия:

специализированные требования по оборудованию отсутствуют; аудитория с посадочными местами по количеству студентов; доска.

6.2. Практические занятия:

1. Проектор;
2. Интерактивная доска.

6.3. Прочее:

1. рабочее место преподавателя, оснащенное компьютером с доступом в Интернет;
2. рабочие места студентов, оснащенные компьютерами с доступом в Интернет, предназначенные для работы в электронной образовательной среде.

Аннотация рабочей программы

Дисциплина **ТЕХНОЛОГИИ КОРПУСНОЙ ЛИНГВИСТИКИ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению *45.05.01 Перевод и переводоведение*. Дисциплина реализуется на факультете *Р* Международного промышленного менеджмента и коммуникации БГТУ "ВОЕНМЕХ" им. Д.Ф. Устинова кафедрой **Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА**.

Дисциплина нацелена на формирование *компетенций*:

ОПК-4 Способен работать с электронными словарями, различными источниками информации, осуществлять поиск, хранение, обработку и анализ информации, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий.

Содержание дисциплины охватывает круг вопросов, связанных с изучением предмета и задач корпусной лингвистики; места корпусной лингвистики в исследовании языка, основных разделов корпусной лингвистики.

Программой дисциплины предусмотрены следующие **виды контроля**:

Текущий контроль успеваемости студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- доклад;
- задание;
- вопросы к зачету.

Промежуточная аттестация проводится в формах:

- экзамен.

Общая трудоемкость освоения дисциплины составляет **3 з.е., 108 ч.** Программой дисциплины предусмотрены лекционные занятия (**17 ч.**), практические занятия (**17 ч.**), самостоятельная работа студента (**74 ч.**).

ТЕХНОЛОГИИ И ФОРМЫ ОБУЧЕНИЯ

Рекомендации по освоению дисциплины для студента

Трудоемкость освоения дисциплины составляет 108 ч., из них 34 ч. аудиторных занятий, и 74 ч., отведенных на самостоятельную работу студента.

Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины приведены в таблице.

Контроль освоения дисциплины производится в соответствии с Положением о текущем, рубежном контроле успеваемости и промежуточной аттестации обучающихся.

Формы контроля и критерии оценивания приведены в приложении 3 к Рабочей программе.

Наименование работы	Рекомендуемая литература	Трудоемкость, час.
Раздел 1. Введение в корпусную лингвистику. Классификация корпусов.		
Изучение темы 1 по конспектам лекций и рекомендованной литературе для подготовки к практическим занятиям. Подготовка докладов.	А. В. Гребенщикова. . Основы квантитативной лингвистики и новых информационных технологий: Москва: Флинта, 2018 (5) А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (3)	20
Итого по разделу 1		20
Раздел 2. История развития корпусной лингвистики. Национальные корпуса текстов.		
Изучение темы 2 по конспектам лекций и рекомендованной литературе для подготовки к практическим занятиям. Подготовка докладов.	А. В. Гребенщикова. . Основы квантитативной лингвистики и новых информационных технологий: Москва: Флинта, 2018 (5) Е. В. Булдакова. . Компьютерная лексикография: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2009 (5)	20
Итого по разделу 2		20
Раздел 3. Стандарты и типы разметки в корпусах.		
Изучение темы 3 по конспектам лекций и рекомендованной литературе для подготовки к практическим занятиям. Подготовка докладов.	А. А. Шунейко. . Корпусная лингвистика: Москва: Юрайт, 2022 (1) А. В. Гребенщикова. . Основы квантитативной лингвистики и новых информационных технологий: Москва: Флинта, 2018 (5)	22
Итого по разделу 3		22
Раздел 4. Сферы практического использования корпусов. Корпусные методы исследований.		
Изучение темы 4 по конспектам лекций и рекомендованной литературе для подготовки к практическим занятиям. Подготовка докладов.	А. А. Шунейко. . Корпусная лингвистика: Москва: Юрайт, 2022 (1) А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (3)	12
Итого по разделу 4		12

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств, позволяющие оценить результаты обучения по данной дисциплине, включают в себя:

- диагностическая работа
- доклад;
- вопросы к зачету;
- задание;
- экзамен.

Критерии оценивания

Диагностическая работа

Диагностическая работа проводится в форме теста в ЭИОС Moodle:

- при правильном ответе менее чем на 60% вопросов - не аттестация;
- при правильном ответе на 60% вопросов и более - аттестация.

Доклад

В семестре предусмотрено выполнение студентами двух докладов.

Оценка «отлично»:

Студент чётко, грамотно и последовательно выстраивает свой ответ; умеет показать связь теории и практики. Раскрывает тему доклада глубоко и всесторонне. Приводит собственные примеры, иллюстрирующие глубокое понимание теоретического материала; исчерпывающе отвечает на вопросы, не допускает неточностей. Доклад чётко структурирован: состоит из введения, основной части и заключения. В заключении подводятся итоги доклада.

Оценка «хорошо»:

Студент грамотно и последовательно излагает материал, умеет показать связь теории и практики, не допускает существенных неточностей в ответах на вопросы, может правильно применять теоретические положения, испытывает незначительные затруднения при самостоятельном обобщении материала. Приводит примеры только из основных источников, не приводит собственных примеров, иллюстрирующих теорию. Раскрывает тему доклада глубоко и всесторонне. Доклад чётко структурирован: состоит из введения, основной части и заключения. В заключении подводятся итоги доклада.

Оценка «удовлетворительно»:

Студент не знает отдельных положений, в ответе допускает неточности, недостаточно правильные формулировки, нарушает последовательность в изложении материала, испытывает затруднения при самостоятельном обобщении материала. Приводит примеры из курса лекций, иллюстрирующие теорию. Доклад не является чётко структурированным.

Оценка «неудовлетворительно»:

Студент не знает значительной части материала, в ответе допускает существенные ошибки, неправильные формулировки, испытывает значительные затруднения при самостоятельном обобщении материала. Не приводит примеров, иллюстрирующих теорию. Доклад не является чётко структурированным.

Вопросы к зачету

Типовой перечень вопросов к зачету (содержит 40 вопросов) представлен в УМК дисциплины.

Задание

Задание направлено на развитие практических навыков работы в корпусными ресурсами (Национальный корпус русского языка, веб-корпуса и пр.): отработка языка корпусных запросов, создание подкорпусов и пр.

Оценка «отлично»:

Студент представил результаты задания, которые корректны и не содержат ошибок.

Оценка «хорошо»:

Студент представил результаты задания, которые корректны и содержат не более 2-х грубых ошибок.

Оценка «удовлетворительно»:

Студент представил результаты задания, которые корректны и содержат не более 4-х грубых ошибок.

Оценка «неудовлетворительно»:
В задании выявлено 5 и более грубых ошибок.

Экзамен

Экзаменационная оценка выставляется в соответствии с баллами, которые студент набрал по технологической карте. В случае несогласия с оценкой студент имеет право сдать экзамен. На экзамене студент отвечает два теоретических вопроса из списка.

Список экзаменационных вопросов.

1. Определение лингвистического корпуса, история корпусов.
2. Требования к разработке корпуса. Основные этапы создания корпуса.
3. Понятия репрезентативности и сбалансированности корпуса.
4. Национальный корпус русского языка: история, состав, виды подкорпусов, система поиска.
5. Британский национальный корпус: история, состав, виды подкорпусов, система поиска.
6. Разметка корпуса, виды разметки.
7. Стандарт TEI в корпусной лингвистике
8. Исследование коллокаций в корпусах.
9. Тематическое моделирование корпусов текстов.
10. Язык корпусных запросов.
11. Корпусные ресурсы устных текстов. Корпус «Один речевой день»: история, состав, система поиска.
12. Параллельные корпусные ресурсы и переводоведение.
13. Веб как корпус.
14. Ресурсы SketchEngine и NoSketchEngine.
15. Выделение ключевых слов в корпусах текстов.
16. Корпус RusProfiling: история, состав, виды подкорпусов, система поиска.
17. Виды метаданных в корпусных ресурсах.
18. Синхронические и диахронические корпусные ресурсы.
19. Связь корпусной лингвистики с другими дисциплинами.
20. Перспективы корпусной лингвистики.

Критерии оценивания

Отлично:

- Полное и точное знание всех основных понятий.
- Умение приводить конкретные примеры.
- Свободное владение методами анализа корпуса.
- Способность аргументированно отвечать на вопросы.
- Умение применять знания на практике.

Хорошо:

- Хорошее знание основных понятий и методов.
- Содержательные ответы с незначительными упущениями.
- Умение применять теоретические знания на практике с некоторой помощью.

Удовлетворительно:

- Базовое понимание ключевых понятий.
- Частично верные ответы с ошибками и неточностями.
- Ограниченное применение знаний на практике.

Неудовлетворительно:

- Недостаточные знания по теме.
- Неполные или неверные ответы.
- Отсутствие понимания основных понятий и методов.
- Неспособность применять знания.

Паспорт фонда оценочных средств

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме			Самостоятельная работа студентов	Формируемая компетенция, %	НАИМЕНОВАНИЕ ОЦЕНОЧНОГО СРЕДСТВА
				ВСЕГО	Лекции	Практические занятия		ОПК-4	
5	9	Раздел 1. Введение в корпусную лингвистику. Классификация корпусов.	30	10	5	5	20	25	Доклад, Вопросы к зачету
5	9	Раздел 2. История развития корпусной лингвистики. Национальные корпуса текстов.	30	10	5	5	20	25	Доклад, Вопросы к зачету
5	9	Раздел 3. Стандарты и типы разметки в корпусах.	28	6	3	3	22	25	Доклад, Вопросы к зачету, Задание
5	9	Раздел 4. Сферы практического использования корпусов. Корпусные методы исследований.	20	8	4	4	12	25	Доклад, Вопросы к зачету, Задание
Всего за 9 семестр			108	34	17	17	74	100	
Всего по дисциплине			108	34	17	17	74	100	

Оценочные материалы по дисциплине ТЕХНОЛОГИИ КОРПУСНОЙ ЛИНГВИСТИКИ

ОПК-4 - Способен работать с электронными словарями, различными источниками информации, осуществлять поиск, хранение, обработку и анализ информации, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий

№ 1 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Какие методы чаще всего используются для обработки многомиллионных корпусных данных?

1. Ручная разметка.
2. Машинное обучение.
3. Разметка с использованием больших языковых моделей.
4. Статистический анализ.

№ 2 Прочитайте текст и запишите развернутый обоснованный ответ

Чем отличается мультимедийный подкорпус НКРЯ от параллельного подкорпуса?

№ 3 Прочитайте текст и запишите развернутый обоснованный ответ

В чем заключаются особенности кореферентной разметки в НКРЯ?

№ 4 Прочитайте текст и установите соответствие

Соотнесите корпус с годом его создания.

Корпус	Год
1. Brown Corpus (Брауновский)	А. 2003
2. British National Corpus (BNC)	Б. 1961
3. Национальный корпус русского языка (НКРЯ)	В. 1994
4. Corpus of Contemporary American English (COCA)	Г. 2008
5. Bank of English (Collins)	Д. 1980-е

№ 5 Прочитайте текст и установите соответствие

Соотнесите корпус с его языком и основным назначением.

Корпус	Характеристика
1. BNC	А. Американский английский, динамический мониторинг языка
2. COCA	Б. Русский язык, включает письменные и устные тексты
3. НКРЯ	В. Британский английский, сбалансированный по жанрам
4. Brown Corpus	Г. Первый электронный корпус (американский английский, 1960-х гг.)
5. Leipzig Corpora Collection	Д. Многоязычные корпуса для исследований и машинного перевода

№ 6 Прочитайте текст и установите последовательность

Расположите этапы корпусного исследования в правильном порядке.

1. Статистический анализ данных.
2. Формулировка исследовательского вопроса.
3. Визуализация результатов.

4. Сбор и обработка текстов.
5. Поиск в корпусе по заданным параметрам.

№ 7 Прочитайте текст и установите последовательность

Расположите следующие технологические достижения в хронологическом порядке их появления.

1. Создание первых POS-тэггеров (частеречных разметчиков).
2. Появление веб-интерфейсов для работы с корпусами.
3. Разработка первых электронных корпусов на магнитных лентах.
4. Внедрение семантической разметки в корпусах.
5. Использование нейросетей для корпусного анализа.

№ 8 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Брауновский корпус был разработан:

1. в начале XXI века
2. в середине XX века
3. в начале XX века
4. в конце XIX века

№ 9 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Репрезентативность и сбалансированность корпуса указывают на:

1. количество и качество
2. теорию и практику
3. семантику и прагматику
4. объем и разметку

№ 10 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Величина ipm указывает на:

1. относительную частоту словоупотребления в корпусе
2. минимальную частоту словоупотребления в корпусе
3. максимальную частоту словоупотребления в корпусе
4. абсолютную частоту словоупотребления в корпусе

№ 11 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Проблема автоматического парсинга Интернет-данных заключается в(о):

1. возможном создании несбалансированного по жанрам корпуса
2. сложности использования библиотек для языка программирования Python
3. вкраплении текстов с разными видами ошибок
4. длительной процедуре постобработки текстов

№ 12 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Ресурс SketchEngine позволяет:

1. строить семантические поля для некоторого слова из корпуса
2. искать коллокации
3. изучать частеречную разметку корпуса
4. хранить рукописные тексты