

УТВЕРЖДАЮ
 Декан факультета

_____ Шматко А.Д.

« ____ » _____ 20__

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ

Направление/специальность подготовки	45.03.03 Фундаментальная и прикладная лингвистика
Специализация/профиль/программа подготовки	Лингвистика в современной информационно-коммуникационной цифровой среде
Уровень высшего образования	Бакалавриат
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	
4	8	3	108	39	26	0	13	69	0	0	69	диф. зач.

ЛИСТ СОГЛАСОВАНИЯ

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

45.03.03 Фундаментальная и прикладная лингвистика

год набора группы: 2025

Программу составил:

Кафедра Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА
Мамаев Иван Дмитриевич, к.ф.н., 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 ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ

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-1 — Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики

ОПК-2 — Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур

Формированию компетенций служит достижение следующих результатов образования:

ОПК-1

знания:

знать основные формальные лингвистические модели;

умения:

уметь оперировать основными формальными лингвистическими моделями;

навыки:

обладать навыком использования математических дисциплин, которые необходимы для формального представления лингвистических знаний.

ОПК-2

знания:

знать основные положения и концепции в области теоретической лингвистики и основного изучаемого языка (языков), теории коммуникации, лингвистического анализа и интерпретации текста;

умения:

уметь обсуждать перспективные направления современной лингвистики;

навыки:

владеть базовыми навыками анализа языковых фактов с использованием традиционных методов и современных информационных технологий.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина **ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению *45.03.03 Фундаментальная и прикладная лингвистика*.

Содержание дисциплины является логическим продолжением дисциплин: **ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ПРОГРАММИРОВАНИЕ, СИНТАКСИС, ОБЩАЯ МОРФОЛОГИЯ, ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ**.

Предварительные компетенции, сформированные у обучающегося до начала изучения дисциплины:

- ОПК-1 — Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики
- ОПК-7 — Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
- ПК-93 — Способен генерировать новые идеи для решения задач цифровой экономики, абстрагироваться от стандартных моделей, перестраивать сложившиеся способы решения задач, выдвигать альтернативные варианты действий с целью выработки новых оптимальных алгоритмов

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-1	ОПК-2
4	8	Раздел 1. Моделирование в естественном языке. Лингвистические модели и их типы. Модели синтезирующие и аналитические. Моделирование различных аспектов речевой деятельности. Модели различных уровней языка и речи (фонетические, морфосинтаксические, семантические). Использование лингвистических моделей в системах автоматической обработки естественного языка.	14	4	4	0	10	20	20
4	8	Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей. Грамматика зависимостей. Конструкции, представляющие трудность для структуры зависимостей. Понятие структуры составляющих. Структурные отношения между составляющими. Сопоставление структуры зависимостей и структуры составляющих. Синтаксическая омонимия: стрелочная омонимия, конституентная омонимия.	18	8	4	4	10	20	20
4	8	Раздел 3. Статистико-комбинаторные модели в лингвистике. Направления компьютерной и прикладной лингвистики, которые используют статистические методы: машинный перевод, автоматическое реферирование, извлечение коллокаций из текстов и др. Статистический аппарат в корпусах текстов: обзор используемых метрик и программных средств.	18	8	4	4	10	20	20
4	8	Раздел 4. Компьютерные модели перевода. История машинного перевода. Три подхода к машинному переводу: rule-based approach, statistical approach, neural approach. Лингвистические проблемы машинного перевода. Понятие постредактирования текстов машинного перевода: требования и стандарты.	14	4	4	0	10	20	20
4	8	Раздел 5. Компьютерные семантические модели. Машинное обучение в лингвистической науке. Тематическое моделирование текстов на естественном языке. Онтологии как способ формального представления лингвистической предметной области. Классификация и кластеризация текстовых данных.	44	15	10	5	29	20	20
Всего за 8 семестр			108	39	26	13	69	100	100
Всего по дисциплине			108	39	26	13	69	100	100

3.2. Аудиторный практикум

№ п/п	Номер и наименование раздела дисциплины	Тема практического занятия	Объем, ауд. часов
1	Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.	Построение грамматик зависимостей и непосредственно составляющих для предложений естественного языка. Анализ систем автоматизированного создания подобных грамматик.	4
2	Раздел 3. Статистико-комбинаторные модели в лингвистике.	Формулы, показывающие меры устойчивости коллокаций. Решение задач.	4
3	Раздел 5. Компьютерные семантические модели.	Сбор корпуса и анализ полученных тематических моделей.	5
Всего за 8 семестр			13

3.3. Самостоятельная работа студента (СРС)

№ п/п	Номер и наименование раздела дисциплины	Содержание учебного задания	Объем, часов
1	Раздел 1. Моделирование в естественном языке.	Самостоятельное изучение формальных моделей в лингвистике по современным научным работам.	10
2	Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.	Ручное и автоматическое создание синтаксических структур отобранных для анализа предложений на естественном языке.	10
3	Раздел 3. Статистико-комбинаторные модели в лингвистике.	Знакомство с методами автоматического извлечения коллокаций из текстов.	10
4	Раздел 4. Компьютерные модели	Работа с современными системами машинного	10

	перевода.	перевода.	
5	Раздел 5. Компьютерные семантические модели.	Создание собственных тематических моделей для отобранных текстовых коллекций.	29
Всего за 8 семестр			69

4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

СЕМЕСТР	НЕДЕЛИ СЕМЕСТРА												
	1	2	3	4	5	6	7	8	9	10	11	12	13
8		Дисск.			ДЗ	ДР	ДЗ		Дисск.	ДР		ДЗ	диф. зач.

Условные обозначения:

- ДР – диагностическая работа;
- Дисск. – дискуссия;
- ДЗ – домашнее задание;
- диф. зач. – дифференцированный зачет.

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

- диагностическая работа;
- дискуссия;
- домашнее задание.

Промежуточная аттестация проводится в формах:

- дифференцированный зачет.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература по дисциплине:

1. А. Н. Гуцин. . Технология обработки текста и звучащей речи. СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018, 66 экз.
2. Д. Ю. Фёдоров. . Программирование на языке высокого уровня Python. Москва: Юрайт, 2023, эл. рес.
3. Э. Алпайдин. . Машинное обучение: новый искусственный интеллект. М.: Альпина Паблишер, 2017, 7 экз.

5.2. Дополнительная литература по дисциплине:

1. А. Н. Баранов. . Введение в прикладную лингвистику. М.: Эдиториал УРСС, 2003, 2 экз.
2. А. Н. Баранов. . Введение в прикладную лингвистику. М.: Эдиториал УРСС, 2001, 1 экз.

5.3. Периодические издания:

не требуются.

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

1. <https://urait.ru/> — Образовательная платформа «Юрайт». Для вузов и ссузов;
2. <https://e.lanbook.com/> — ЭБС Лань;
3. <https://scholar.google.com/> — Google Академия;
4. <http://library.voenmeh.ru/jirbis2> — Сайт фундаментальной библиотеки БГТУ «Военмех» им. Д.Ф. Устинова — Фундаментальная библиотека БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова;
5. <http://stanza.run/> — Stanza;
6. <https://mimno.infosci.cornell.edu/jsLDA/> — jsLDA: In-browser topic modeling.

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 **ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению 45.03.03 *Фундаментальная и прикладная лингвистика*. 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-1 Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики;

ОПК-2 Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур.

Содержание дисциплины охватывает круг вопросов, связанных с современными исследованиями в области прикладной, математической и компьютерной лингвистики. Курс содержит начальные сведения о формальных моделях, формальной семантике, истории структурной лингвистики и компьютерной лингвистики.

Программой дисциплины предусмотрены следующие **виды контроля**:

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

- диагностическая работа;
- дискуссия;
- домашнее задание.

Промежуточная аттестация проводится в формах:

- дифференцированный зачет.

Общая трудоемкость освоения дисциплины составляет 3 з.е., **108 ч**. Программой дисциплины предусмотрены лекционные занятия (**26 ч.**), практические занятия (**13 ч.**), самостоятельная работа студента (**69 ч**).

ТЕХНОЛОГИИ И ФОРМЫ ОБУЧЕНИЯ

Рекомендации по освоению дисциплины для студента

Трудоемкость освоения дисциплины составляет 108 ч., из них 39 ч. аудиторных занятий, и 69 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. Моделирование в естественном языке.		
Самостоятельное изучение формальных моделей в лингвистике по современным научным работам.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (1)	10
Итого по разделу 1		10
Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.		
Ручное и автоматическое создание синтаксических структур отобранных для анализа предложений на естественном языке.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (4)	10
Итого по разделу 2		10
Раздел 3. Статистико-комбинаторные модели в лингвистике.		
Знакомство с методами автоматического извлечения коллокаций из текстов.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (2) Д. Ю. Фёдоров. . Программирование на языке высокого уровня Python: Москва: Юрайт, 2023 (1)	10
Итого по разделу 3		10
Раздел 4. Компьютерные модели перевода.		
Работа с современными системами машинного перевода.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2001 (4) А. Н. Гуцин. . Технология обработки текста и звучащей речи: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018 (17)	10
Итого по разделу 4		10
Раздел 5. Компьютерные семантические модели.		
Создание собственных тематических моделей для отобранных текстовых коллекций.	Э. Алпайдин. . Машинное обучение: новый искусственный интеллект: М.: Альпина Паблишер, 2017 (1–3)	29
Итого по разделу 5		29

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

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

- диагностическая работа
- дискуссия;
- домашнее задание;
- дифференцированный зачет.

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.

Дискуссия

Во время дискуссии студенты и преподаватель обсуждают современные проблемы в области формальных лингвистических моделей. Приблизительный список дискуссионных тем приведён ниже.

1. Синтаксические конструкции зависимостей.
2. Синтаксические конструкции составляющих.
3. Подходы к машинному переводу.
4. Статистические меры анализа устойчивости коллокаций.
5. Семантические модели.

Если во время дискуссии обучающийся всестороннее понимания вопроса, то такое обсуждение может быть зачтено. Если преподаватель не удовлетворён ответом, он может задать студентам дополнительные дискуссионные вопросы.

Домашнее задание

Домашнее задание обсуждается с преподавателем перед окончанием обсуждения определённого тематического блока лекционно-практического курса. Домашнее задание может быть оформлено в виде отчёта как в рукописном, так и в печатном формате. При оценке отчёта преподаватель учитывает глубину раскрытия вопроса, а также верность хода решения. Если домашнее задание включало обработку текстовых массивов, то в отчёте также необходимо предоставить листинг программы для обработки или ссылку на используемый для обработки ресурс.

Если обучающийся получил верный ответ, подробно объяснил процесс решения задачи и обосновал выбор того или иного подхода, то домашняя работа может быть зачтена (5 баллов). Наличие не более 2 серьёзных недочётов приводит к снижению общего балла до 4. Если студент допустил от 2 до 4 недочётов, то максимальный общий балл не превышает 3. Если обучающийся неверно решил задачу или же показал поверхностные знания при объяснении процесса решения, то ему рекомендуется решить задачу ещё раз после комментариев преподавателя.

Дифференцированный зачет

Дифференцированный зачет выставляется согласно технологической карте, которая составляется с учетом приказа ректора №687-О от 08.09.2023.

Если студент не согласен с предлагаемой оценкой, то он имеет право сдать дифференцированный зачёт в устной форме по вопросам, которые обсуждались в цикле всего курса, а именно: общие положения о формальных моделях в лингвистике, синтаксические формальные модели, модели машинного перевода, статистико-комбинаторные модели в лингвистике, лингвистические онтологии, тематическое моделирование.

Оценка «зачтено-отлично» выставляется обучающемуся, если он показывает всесторонние, систематизированные, глубокие знания вопросов по темам дисциплины, а также умение уверенно применять их на практике.

Оценка «зачтено-хорошо» выставляется обучающемуся, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе некоторые неточности, которые может устранить с помощью дополнительных вопросов преподавателя.

Оценка «зачтено-удовлетворительно» выставляется обучающемуся, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными понятиями выносимых на опрос тем, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации.

Оценка «не зачтено» выставляется обучающемуся, который не знает большей части основного содержания вопросов тем дисциплины, допускает грубые ошибки в формулировках основных понятий и не умеет использовать на практике полученные 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-1	ОПК-2	
4	8	Раздел 1. Моделирование в естественном языке.	14	4	4	0	10	20	20	Дискуссия
4	8	Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.	18	8	4	4	10	20	20	Домашнее задание
4	8	Раздел 3. Статистико-комбинаторные модели в лингвистике.	18	8	4	4	10	20	20	Домашнее задание
4	8	Раздел 4. Компьютерные модели перевода.	14	4	4	0	10	20	20	Дискуссия
4	8	Раздел 5. Компьютерные семантические модели.	44	15	10	5	29	20	20	Домашнее задание
Всего за 8 семестр			108	39	26	13	69	100	100	
Всего по дисциплине			108	39	26	13	69	100	100	

ОПК-1 - Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики

- № 1 Прочитайте текст и установите соответствие
Соотнесите виды языковых моделей с их описанием.

Модель	Описание
1. Униграммная модель.	А. Векторные представления слов.
2. N-граммная модель.	Б. Аппроксимации вероятности на основе предыдущего контекста.
3. Нейросетевая модель.	В. Комбинация нескольких конечных автоматов с одним состоянием.

- № 2 Прочитайте текст и установите соответствие
Сопоставьте элементы формальной онтологии с их значением.

Элемент	Значение
1. At	А. Атрибуты онтологии.
2. E	Б. Экземпляры онтологии.
3. A	В. Отношения онтологии.
4. R	Г. Аксиомы онтологии.

- № 3 Прочитайте текст и установите последовательность
Расположите этапы тематического моделирования в правильном порядке.

1. Визуализация результатов.
2. Предобработка текстов.
3. Интерпретация тематических кластеров.
4. Обучение модели.
5. Оценка качества модели.

- № 4 Прочитайте текст и установите последовательность
Расположите этапы работы BERTopic в правильном порядке.

1. Кластеризация эмбедингов (UMAP + HDBSCAN).
2. Векторизация текстов (создание эмбедингов).
3. Извлечение ключевых слов для тем (с помощью c-TF-IDF).
4. Постобработка и объединение схожих тем.
5. Визуализация тематических кластеров.

- № 5 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа
Что такое машинный перевод?

1. Процесс перевода речи с помощью человека.
2. Автоматизированный процесс перевода текста или речи с одного языка на другой.
3. Техника передачи информации через Интернет.
4. Процесс анализа грамматической структуры предложений.

- № 6 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа
Что представляет собой модель *encoder-decoder* в контексте машинного перевода?

1. Модель, использующая только грамматические правила.
2. Система, основанная на статистических методах.
3. Архитектура нейронной сети для перевода.

4. Модель, работающая только с текстом на одном языке.

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

И.А. Мельчук разработал модель «_____ ↔ Текст».

1. Морфема
2. Семантика
3. Смысл
4. Прагматика

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

Выберите верные аналоги отношений категоризации в онтологии.

1. IS-A.
2. PART-OF.
3. A-KIND-OF.
4. Родовидовые отношения.

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

Библиотека NLTK может реализовать один из следующих синтаксических формализмов:

1. грамматика зависимостей
2. грамматика непосредственно составляющих
3. грамматика непосредственных зависимостей
4. грамматика непосредственностей

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

Выберите приложения, в которых возможно автоматически вывести синтаксическую структуру предложения.

1. AOT
2. LinkGrammar
3. OpenCorpora
4. ЭТАП
5. mystem

№ 11 Прочитайте текст и запишите развернутый обоснованный ответ

Что такое онтология в общем смысле?

№ 12 Прочитайте текст и запишите развернутый обоснованный ответ

Какие три типа формальных лингвистических моделей выделяет Ю. Д. Апресян в своих работах?

ОПК-2 - Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур

№ 1 Прочитайте текст и запишите развернутый обоснованный ответ

Когда применяется предиктивный (предсказывающий) морфологический анализ?

№ 2 Прочитайте текст и запишите развернутый обоснованный ответ

Что представляет собой метод *Word Embeddings*? Какие у нее особенности?

№ 3 Прочитайте текст и установите соответствие

Установите соответствие между автоматически получаемыми морфологическими характеристиками в библиотеке *rumorphy2* (из базы данных *OpenCorpora*) с их значением.

Морфологическая характеристика	Значение
--------------------------------	----------

- | | |
|---------|----------------------|
| 1. inan | А. несовершенный вид |
| 2. tran | Б. переходность |

- | | |
|---------|-----------------------------|
| 3. indc | В. неодушевленность |
| 4. impf | Г. изъявительное наклонение |

№ 4 Прочитайте текст и установите соответствие
Соотнесите большие языковые модели (LLM) с их особенностями.

Модель	Характеристика
1. GPT-3	А. Первая модель с трансформерной архитектурой (2017).
2. BERT	Б. Использует только декодер, генерация текста по промту.
3. T5	В. Заточена под задачи перевода и классификации, использует encoder-decoder.
4. Transformer	Г. Двухнаправленная модель, предобучена на заполнении пропусков (masked language model).

№ 5 Прочитайте текст и установите последовательность
Расположите этапы в правильном порядке.

1. Предобработка текстов.
2. Векторизация текста.
3. Визуализация и интерпретация кластеров.
4. Применение алгоритма кластеризации.
5. Выбор метрики расстояния.

№ 6 Прочитайте текст и установите последовательность
Расставьте этапы лингвистического эксперимента в порядке его выполнения.

1. Собрать лингвистические данные.
2. Обработать лингвистические данные.
3. Подтвердить или опровергнуть гипотезу.
4. Найти лингвистическую проблему или лагуну.

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

В формализме Universal Dependencies для обозначения прямых дополнений используется следующий тег:

1. subj
2. indobj
3. proobj
4. obj

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

Какая технология позволяет генерировать текст, имитирующий стиль и содержание обучающего корпуса?

1. Датасет.
2. Языковая модель.
3. Адаптационная модель.
4. Речевой аппарат.

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

В формализме Universal Dependencies тег nsubj обозначает:

1. no subject
2. normal subject
3. new subject
4. nominal subject

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

Какие онтологии существуют с точки зрения языка?

1. Одноязычные.
2. Верхней зоны.
3. Многоязычные.
4. Средней зоны.

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

Какие виды синтаксической омонимии предложены А.В. Гладким?

1. Стрелочная омонимия.
2. Разметочная омонимия.
3. Кластерная омонимия.
4. Подчинительная омонимия.

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

Выберите виды вероятностных тематических моделей.

1. LSA.
2. pLSA.
3. LDA.
4. VSM.