

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Балашов Александр Геннадьевич
Должность: Проректор по учебной работе
Дата подписания: 25.02.2026 14:43:31
Уникальный программный ключ:
159a23a86061c60c9ea58b6d4823a92e747fed78

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Московский институт электронной техники»

УТВЕРЖДЕНА

Проректор по УР



А.Г. Балашов

2025 г.

Дополнительная профессиональная программа
(программа профессиональной переподготовки)

Компьютерная лингвистика

(наименование программы)

Москва 2025

Аннотация

Дополнительная профессиональная программа профессиональной переподготовки ИТ-профиля (далее – Программа) предназначена для обучающихся по очной или очно-заочной форме за счет бюджетных средств или по договорам об оказании платных образовательных услуг, освоивших программы:

- 1) *бакалавриата в объеме не менее 1 курса (бакалавры 2 курса), специалитета в объеме не менее 1 курса (специалисты 2 курса) и программы магистратуры (магистры) по специальностям и направлениям подготовки Лингвистика, Филология, Фундаментальная и прикладная лингвистика, Интеллектуальные системы в гуманитарной сфере, Перевод и переводоведение и другим направлениям подготовки в приоритетной отрасли «социальная сфера» (ЦГ1)..*

Целью профессиональной переподготовки является получение актуальной для социальной отрасли дополнительной ИТ-квалификации: специалист по обработке естественного языка.

Нормативный срок освоения программы 470 часов при очно-заочной форме подготовки с применением электронного обучения и дистанционных образовательных технологий).

Авторы и преподаватели:

№ п/п	ФИО, должность	Модули	Часов, всего
1	Коночкин К.Н., программист АО НТЦ "ЭЛИНС"	ВМ на Python	42
2	Капитанов А.И., к.т.н., разработчик отдела бэкенд-разработки ООО «ГринСайт»	Python для анализа данных	66
3	Ефремова М. П., к.фил.н., доцент СПбГЭУ	Структурная лингвистика	66
4	Грабарник К.Л., зам. дир. НИУ МИЭТ		
5	Ефремова М. П., к.фил.н., доцент СПбГЭУ	Функциональные и когнитивные модели в лингвистике	70
6	Бордюжа Виктор, глава направления	SQL и NoSQL базы данных	50

	компьютерного зрения и машинного обучения ООО "ВИЛЛ НАЙН"		
7	Ефремова Мария Павловна, доцент СПбГЭУ	Автоматическая обработка текста	72
8	Бордюжа Виктор, глава направления компьютерного зрения и машинного обучения ООО "ВИЛЛ НАЙН"	Нейронные сети для обработки естественного языка	66
9	Капитанов А.И., к.т.н., разработчик отдела бэкенд-разработки ООО «ГринСайт»		
10	Фирсанова В. И., преподаватель НИУ ВШЭ		

Подробная информация о профессорско-преподавательском составе (ФИО, должность, степень, опыт работы) представлена отдельным документом (загружена и хранится на облачном диске ФГАНУ «Социоцентр»).

I. Общие положения

I.1. Нормативная правовая основа Программы

Дополнительная профессиональная программа (программа профессиональной переподготовки) ИТ-профиля «Компьютерная лингвистика» (далее – Программа) разработана в соответствии со следующей нормативно-правовой основой:

- Федеральный закон от 29 декабря 2012 г. № 273-ФЗ «Об образовании в Российской Федерации»;
- постановление Правительства РФ от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;
- приказ Минобрнауки России от 24 марта 2025 г. № 266 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника, утвержденный приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. N 929. (далее вместе – ФГОС ВО);
- проект Приказа Минтруда России "Об утверждении профессионального стандарта "Специалист по машинному обучению" (по состоянию на 03.06.2020) (подготовлен Минтрудом России, ID проекта 01/02/06-20/00102579).

I.2. Термины и определения, используемые в Программе

Итоговая аттестация (аттестация) – оценка степени и уровня освоения обучающимися ДПП ПП в формате демонстрационного экзамена, предусматривающая выполнение обучающимся профессиональных задач и оценку результатов и (или) процесса выполнения, проверку сформированности цифровых компетенций в ходе обучения по ДПП ПП.

Демонстрационный экзамен – аттестационное испытание, предусматривающее выполнение профессиональных задач и оценку результатов и (или) процесса выполнения профессиональных задач для подтверждения применения

обучающимися цифровых компетенций на практике.

Дистанционные образовательные технологии – это образовательные технологии, реализуемые в основном с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) взаимодействии обучающихся и педагогических работников.

Дополнительная профессиональная программа профессиональной переподготовки (Программа) – комплекс основных характеристик образования (объем, содержание, планируемые результаты) и организационно-педагогических условий, который представлен в виде учебного плана, календарного учебного графика, рабочих программ учебных курсов, дисциплин (модулей), оценочных и методических материалов, а также программ учебной и производственной практик, стажировок и форм аттестации, иных компонентов и обеспечивает приобретение дополнительной квалификации. Программа может разрабатываться с учетом положений профессиональных стандартов, федеральных государственных образовательных стандартов, требований рынка труда (индустрии).

Знание (З) – информация о свойствах объектов, закономерностях процессов и явлений, правилах использования этой информации для принятия решений, присвоенная обучающимся на одном из уровней, позволяющих выполнять над ней мыслительные операции.

Матрица компетенций – матрица компетенций, актуальных для цифровой экономики, с приоритетом компетенций в ИТ-сфере, разработанная Университетом Иннополис при участии ИТ-компаний и университетов-участников программы «Приоритет-2030», представляющая собой перечень компетенций, структурированный по сферам применения, типу компетенций, уровням их сформированности и характеристикам.

Опыт практической деятельности (ОПД) – образовательный результат, включающий выполнение обучающимся деятельности, завершающейся получением результата / продукта (элемента продукта), значимого при выполнении трудовой функции, в условиях реального производства или в модельной ситуации.

Оценочные средства (ОС) – дидактические средства для оценки качества

подготовленности обучающихся.

Практика (практическая подготовка) – форма организации образовательной деятельности при освоении образовательной программы в условиях выполнения обучающимися определенных видов работ, связанных с будущей профессиональной деятельностью и направленных на формирование, закрепление, развитие практических навыков и компетенции по профилю соответствующей образовательной программы.

Профессиональный модуль (ПМ) – структурный элемент Программы, предназначенный для формирования определенной компетенции или нескольких компетенций.

Рабочая программа – нормативный документ в составе Программы, регламентирующий взаимодействие преподавателя и обучающихся в ходе учебного процесса при реализации структурных элементов Программы (модуль, дисциплина, курс).

Умение (У) – освоенный субъектом способ выполнения действия, обеспечиваемый совокупностью приобретенных знаний и навыков; операция (действие), выполняемая определенным способом и с определенным качеством (умение, выполнение которого доведено до автоматизма, является навыком).

Учебная дисциплина (УД) – структурный элемент Программы, предназначенный для формирования знаний и умений в соответствующей сфере профессиональной деятельности.

Фонды оценочных средств (ФОС) – совокупность оценочных средств, используемых на различных этапах педагогической диагностики.

Целевой уровень сформированности компетенции – определенный в соответствии с Матрицей цифровых компетенций и указанный в ДПП ПП в качестве планируемого результата обучения уровень сформированности цифровой компетенции.

Цифровая компетенция (компетенция) – образовательный результат, формируемый при освоении ДПП ПП и необходимый для приобретения дополнительной *ИТ-квалификации*, необходимой для выполнения нового вида деятельности по внедрению и (или) развитию, и (или) разработке цифровых технологий, в том числе

алгоритмов и программ, пригодных для практического применения, в одной из приоритетных отраслей экономики.

Электронное обучение – организация образовательной деятельности с применением содержащейся в базах данных и используемой при реализации образовательных программ информации и обеспечивающих ее обработку информационных технологий, технических средств, а также информационно-телекоммуникационных сетей, обеспечивающих передачу по линиям связи указанной информации, взаимодействие обучающихся и педагогических работников.

1.3. Требования к поступающим

К обучению по Программе допускаются обучающиеся очной, очно-заочной форме за счет бюджетных средств или по договорам об оказании платных образовательных услуг, освоившие программы ВО *бакалавриата в объеме не менее 1 курса (бакалавры 2 курса), специалитета в объеме не менее 1 курса (специалисты 2 курса) и программы магистратуры (магистры) по специальностям и направлениям подготовки Лингвистика, Филология, Фундаментальная и прикладная лингвистика, Интеллектуальные системы в гуманитарной сфере, Перевод и переводоведение и другим направлениям подготовки в приоритетной отрасли «социальная сфера» (ЦГ1).*

1.4. Квалификационная характеристика выпускника

Выпускникам Программы присваивается дополнительная ИТ-квалификация в области обработки естественного языка в социальной сфере (лингвистика).

Выпускник Программы будет готов к выполнению трудовой деятельности в **области создания прикладных информационных систем с использованием методов машинного обучения и интеллектуальной обработки данных в качестве специалиста по обработке естественного языка (инженера по обработке естественного языка),** и выполнению трудовых функций в области **разработки моделей машинного обучения для обработки естественного языка и адаптации моделей машинного обучения для разработки прикладных решений с использованием средств обработки естественного языка.**

Отнесение к видам экономической деятельности:

62.01	Разработка компьютерного программного обеспечения
(код ОКВЭД)	(наименование вида экономической деятельности)

Квалификационный уровень по национальной рамке квалификаций: 6.

II. Планируемые результаты обучения и структура Программы

Получение дополнительной ИТ-квалификации «специалист по обработке естественного языка» обеспечивается формированием приведённых в таблице цифровых компетенций:

Наименование сферы	ID и наименование компетенции	Инструменты профессиональной деятельности	Целевой уровень формирования компетенций в Программе			
			Минимальный (исходный)	Базовый	Продвинутый	Экспертный
Средства программной разработки	28 Применяет языки программирования для решения профессиональных задач	Python			Самостоятельно применяет языки программирования и настраиваемые программные инструменты для автоматизации процессов в профессиональной деятельности	
Искусственный интеллект и машинное обучение	37 Применяет Искусственный интеллект и машинное обучение	Mystem, Pymorphy, NLTK, SpaCy, Natasha, PyTorch или TensorFlow HuggingFace			Разрабатывает отдельные части проектов по применению искусственного интеллекта и машинного обучения	
Основы инженерно-научных расчетов на Python	15 Применяет программные алгоритмы обработки данных для инженерно-научных	Python			Применяет программные алгоритмы обработки данных для инженерно-научных расчетов, самостоятельно подбирает и использует	

Наименование сферы	ID и наименование компетенции	Инструменты профессиональной деятельности	Целевой уровень формирования компетенций в Программе			
			Минимальный (исходный)	Базовый	Продвинутый	Экспертный
	расчетов				для научных вычислений специализированные библиотеки. Разрабатывает программные алгоритмы с использованием методов математической оптимизации	
Средства программной разработки	32Использует СУБД при разработке ПО	Tantor		Знает основы баз данных, знаком с нормализацией, ACID, транзакциями, может написать простые выборки («select'ы»). Участвует в проектах по созданию ПО с использованием СУБД под контролем опытных специалистов		

II.1. Структура образовательных результатов

Формирование цифровых компетенций, необходимых для получения обучающимися дополнительной ИТ-квалификации, обеспечивается последовательным формированием промежуточных образовательных результатов, начиная со знаний.

ID и формулировка целевого уровня формирования компетенций	Промежуточные образовательные результаты		
	Опыт практической деятельности (ОПД)	Умения (У)	Знания (З)
ID 28 Самостоятельно применяет языки программирования и настраиваемые программные инструменты для автоматизации процессов в профессиональной деятельности	ОПД1 использования синтаксиса и библиотек Python/SQL/других языков в соответствии с задачами; интеграции отдельных программных модулей в единое приложение, оформление кода с комментариями.	У1 применять современные методы компьютерной реализации геометрических и алгебраических моделей к решению задач; создавать отчеты о проделанной работе с пояснениями, графиками и выводами.	З1 понятий и основных технических приемов матричной алгебры, аналитической геометрии, булевой алгебры, теории линейных операторов и квадратичных форм; элементов теории кодирования.
		У2 писать программы на Python для анализа данных; использовать библиотеки Python для обработки и анализа информации; проектировать и реализовывать классы и модули; отлаживать и тестировать код; применять многопоточность и создавать сетевые приложения.	З2 синтаксиса Python и основных конструкций; методов работы с файлами и функциями; принципов объектно-ориентированного программирования; методов обработки ошибок; основ многопоточности и сетевого взаимодействия.
ID 37 Разрабатывает отдельные части проектов по применению искусственного интеллекта и машинного обучения	ОПД2 выбора готовой модели ИНС (искусственной нейронной сети) или разработки собственной, обработки данных для обучения/тестирования модели, оценка качества работы модели по метрикам.	У3 применять библиотеки для обработки текстов и работы с нейронными сетями; разрабатывать и оптимизировать пайплайны для решения задач в сфере обработки естественного языка; работать с API больших языковых моделей; разрабатывать программные модули для обработки естественного языка.	З3 современных методов обработки естественного языка; алгоритмов построения векторных представлений текста; методов предобработки текстовых данных; принципов работы нейронных сетей для обработки естественного языка; основ работы с большими языковыми моделями.

		У4 проводить фонетический и морфологический анализ слов и предложений; строить деревья зависимостей и применять методы генеративной грамматики; использовать библиотеки для NLP для предобработки текста и построения моделей классификации и регрессии; разрабатывать и оценивать простые языковые модели; анализировать результаты машинного обучения, корректно интерпретировать метрики.	34 основных понятий и теорий фонетики и фонологии, включая глубинные представления, генеративную фонологию и теорию оптимальности; строения морфологической системы языка и методы морфологического анализа; принципов генеративной и X-штрих грамматики; семантических типов, рестрикторов и кванторов и их роль в структуре предложения; основных подходов в обработке естественного языка и вероятностных моделей.
ID 15 Применяет программные алгоритмы обработки данных для инженерно-научных расчетов, самостоятельно подбирает и использует для научных вычислений специализированные библиотеки. Разрабатывает программные алгоритмы с использованием методов математической оптимизации	ОПД3 анализа входных данных и выбора алгоритма их обработки, реализации алгоритма в виде работающего кода, проведения тестирования и оптимизации решения.	У5 выполнять предобработку текста; реализовывать методы rule-based и статистического NER, POS-тегирования и стемминга/лемматизации на Python; строить конечные автоматы для морфологических задач; получать данные через API и web-скрейпинг, соблюдать этические нормы; реализовывать простейшие парсеры на базе CFG и алгоритмы переходного-ориентированного анализа; применять методы тематического моделирования и вычислять меры сходства; строить классификаторы и проводить анализ тональности; реализовывать экстрактивные алгоритмы суммаризации и простые QA-системы; работать с многоязычными данными и извлекать ключевые слова; выполнять сетевой анализ текстов и оценивать сходство текстов. У6 применять принципы функциональной и когнитивной лингвистики при анализе языкового материала; использовать корпусные методы исследования для решения поставленных лингвистических задач;	35 основных этапов обработки текстов; методов распознавания именованных сущностей и оценки качества; методов морфологического анализа и POS-тегирования, различий между стеммингом и лемматизацией; основ конечных автоматов (FSA/FST) для морфологических задач; принципов работы с API и сбора данных из открытых источников; контекстно-свободных и зависимых грамматик, алгоритмов CYK и Earley, методов извлечения коллокаций; тематического моделирования (LDA) и мер сходства текстов; методов классификации, суммаризации и анализа тональности текстов; особенностей обработки многоязычных текстов и алгоритмов извлечения ключевых слов; основ анализа текстов социальных медиа и мер сходства текстов. 36 основных принципов функционализма; теории фреймов Ч. Филлмора и её применение в лингвистике; основных моделей грамматики конструкций (Филлмора, Лакоффа-Гольдберга, Крофта);

		работать с лингвистическими ресурсами (FrameNet, НКРЯ) для извлечения и систематизации данных; разрабатывать простые аналитические алгоритмы для обработки языковых данных с использованием стандартных библиотек; применять модели грамматики конструкций и фреймовой семантики для интерпретации и описания языковых явлений.	технологии работы с лингвистическими корпусами; корпусных методов исследований; основных лингвистических ресурсов и баз данных (FrameNet, НКРЯ).
ID 39 Использует СУБД при разработке ПО Знает основы баз данных, знаком с нормализацией, ACID, транзакциями, может написать простые выборки («select'ы»). Участвует в проектах по созданию ПО с использованием СУБД под контролем опытных специалистов	ОПД4 разработки и оптимизации запросов; работы с реляционными и NoSQL СУБД; взаимодействия СУБД с внешними приложениями; проектирования структур данных.	У7 проектировать и создавать структуры баз данных; формировать SQL-запросы различной сложности; выполнять операции выборки, обновления и удаления данных; интегрировать СУБД с клиентскими приложениями; использовать NoSQL базы данных для решения прикладных задач.	37 классификации и архитектуры СУБД; основных моделей данных и модель ANSI/SPARC; основ реляционной алгебры; синтаксиса SQL и его основных конструкций; принципов работы NoSQL баз данных.

II.2. Структура Программы

Структура Программы регулирует образовательные траектории обучающихся, последовательность освоения структурных элементов (разделов) Программы, соответственно, последовательность формирования всех образовательных результатов.

Структурные элементы (разделы Программы)	Шифры образовательных результатов	Вариатив / инвариант и целевые группы обучающихся
Модуль 1. Высшая математика на Python	Знания: З1 Умения: У1	ЦГ1
Модуль 2. Python для анализа данных	Знания: З2 Умения: У2	ЦГ1
Модуль 3. Структурная лингвистика	Знания: З4 Умения: У4	ЦГ1
Модуль 4. Функциональные и когнитивные модели в лингвистике	Знания: З6 Умения: У6	ЦГ1
Модуль 5. SQL и NoSQL базы данных	Знания: З7 Умения: У7	ЦГ1
Модуль 6. Автоматическая обработка текста	Знания: З5 Умения: У5	ЦГ1
Модуль 7. Нейронные сети для обработки естественного языка	Знания: З3 Умения: У3	ЦГ1
Практика	Опыт практической деятельности: ОПД1, ОПД2, ОПД3, ОПД4	ЦГ1

III. Учебный план Программы

Объем Программы составляет **470** часов.

Учебный план Программы определяет перечень, последовательность, общую трудоемкость разделов и формы контроля знаний.

Структурные элементы (разделы Программы)	Общая трудоемкость, часов	Обязательная аудиторная учебная нагрузка		Самостоятельная работа, часов		Практики, стажировки, часов	Промежуточная аттестация, часов
		всего, часов	в т.ч. практические занятия, часов	всего, часов	в т.ч. практические задания, часов		
Модуль 1. Высшая математика на Python	42	32	16	8	5		2
Модуль 2. Python для анализа данных	66	48	32	16	8		2
Модуль 3. Структурная лингвистика	66	48	32	16	8		2
Модуль 4. Функциональные и когнитивные модели в лингвистике	70	48	32	20	12		2
Модуль 5. SQL и NoSQL базы данных	50	28	20	20	10		2
Модуль 6. Автоматическая обработка текста	72	42	28	28	14		2
Модуль 7. Нейронные сети для обработки естественного языка	66	48	32	16	8		2
Практика	32					32	
Итоговая аттестация в формате демонстрационного экзамена (включая подготовку к аттестации)				6			
Итого:	470	294	192	130		32	14

IV. Календарный учебный график

Календарный учебный график представляет собой график учебного процесса, рассчитанный на 39 учебных недель и устанавливающий последовательность и продолжительность обучения, включая практику и итоговую аттестацию, а также этапы оценки цифровых компетенций. При этом время, выделяемое на прохождение оценки сформированности цифровых компетенций, в общей трудоемкости Программы, отраженной в Учебном плане, не учитывается.

Структурные элементы (разделы Программы)	Учебные недели			
	1 - 18	19- 36	37	38-39
Модуль 1. Высшая математика на Python				
Модуль 2. Python для анализа данных				
Модуль 3. Структурная лингвистика				
Модуль 4. Функциональные и когнитивные модели в лингвистике				
Модуль 5. SQL и NoSQL базы данных				
Модуль 6. Автоматическая обработка текста				
Модуль 7. Нейронные сети для обработки естественного языка				
Практика				
Аттестация в формате демонстрационного экзамена				

V. Рабочие программы модулей (курсов, дисциплин)

Рабочие программы разрабатываются для структурных элементов (разделов) Программы, указанных в Структуре Программы и Учебном плане, и содержат: перечень тем, включающих лекции, семинары, мастер-классы, практические занятия, самостоятельную работу, консультации и иные виды учебной работы с указанием краткого содержания и трудоемкости, образцы оценочных средств, методические материалы для преподавателей и обучающихся, сведения о кадровом обеспечении образовательного процесса.

Рабочая программа практики / стажировки предусматривает определение цели и задач практической деятельности обучающихся, площадку (площадки) прохождения практики, задания (индивидуальные или групповые), критерии оценки результатов практической деятельности обучающихся.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

1. Рабочая программа модуля

Высшая математика на языке Python

(наименование модуля)

Москва 2025

1.1 Область применения рабочей программы

Рабочая программа модуля «Высшая математика на языке Python» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Компьютерная лингвистика» и направлена на формирование студентами знаний и умений в области решения профессиональных задач на Python.

Рабочая программа нацелена на получение следующей компетенции:

ID 28 Применяет языки программирования для решения профессиональных задач на продвинутом уровне: самостоятельно применяет языки программирования и настраиваемые программные инструменты для автоматизации процессов в профессиональной деятельности.

Индикаторы достижения компетенции:

Знает (З1) понятия и основные технические приемы матричной алгебры, аналитической геометрии, булевой алгебры, теории линейных операторов и квадратичных форм; элементы теории кодирования.

Умеет (У1) применять современные методы компьютерной реализации геометрических и алгебраических моделей к решению задач; создавать отчеты о проделанной работе с пояснениями, графиками и выводами.

Программа является обязательной к освоению.

1.2 Структура и краткое содержание программы

№ п/ п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1	Тема 1. Элементы линейной алгебры Лекция 1-3. Элементы линейной алгебры. Линейное пространство двоичных векторов. Матрицы. Операции над матрицами. Обратная матрица. Решение систем линейных уравнений. Поле $\mathbb{Z}_2$. Решение систем линейных уравнений над полем $\mathbb{Z}_2$. Лабораторная работа 1-3. Построение графиков. Визуализация векторов. Работа с векторами в Python. Представление и вывод матриц в Python. Работа с матрицами в Python. Операции над матрицами в Python. Составление и решение систем линейных уравнений в Python. Работа с полем $\mathbb{Z}_2$ в Python. Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий 1-3.	6 (лекции) 6 (лабораторные) 3 (СРС)

2	Тема 2. Алгебра логики Лекция 4-6. Алгебра логики. Высказывания и логические операции над ними. Функции алгебры логики. Формулы алгебры логики. СДНФ и СКНФ. Равносильные преобразования формул. Бинарные отношения на множестве. Свойства бинарных отношений. Отношения эквивалентности и порядка. Частично упорядоченные множества. Лабораторная работа 4-6. Логические операторы в Python. Составление и вывод таблиц истинности в Python. Модули библиотеки sympy для решения задач булевой алгебры. Составление и анализ СДНФ, СКНФ и полинома Жегалкина в Python. Работа со множествами и бинарными отношениями в Python. Иллюстрация бинарных отношений с помощью матриц, графиков и орграфов в Python. Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий 4-6.	6 (лекции) 6 (лабораторные) 3 (СРС)
3	Тема 3. Элементы теории кодирования Лекция 7-8. Элементы теории кодирования. Линейные коды. Алгоритм кодирования Шеннона – Фано. Коды Хэмминга. Лабораторная работа 7-8. Кодирование текстовых данных и изображений в Python. Алгоритм кодирования Шеннона – Фано в Python. Алгоритм кодирования Хэмминга в Python. Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий 7-8.	4 (лекции) 4 (лабораторные) 2 (СРС)

4	Промежуточная аттестация в формате тестирования и решения практического задания	2
---	--	---

1.3 Учебно-тематический план программы

№ п/ п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов			
		аудиторных		самостоятельной работы	
		всего, часов	практические занятия	всего, часов	практические занятия
1	Тема 1. Элементы линейной алгебры	12	6	1	2
2	Тема 2. Алгебра логики	12	6	1	2
3	Тема 3. Элементы теории кодирования	8	4	1	1
	Промежуточная аттестация	2			
	Итого	42			

1.4 Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение текущего, промежуточного контроля. Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Для текущего и промежуточного контроля создаются фонды оценочных средств (ФОС). ФОС включают в себя педагогические контрольно-измерительные

материалы, предназначенные для определения соответствия (или несоответствия) индивидуальных образовательных достижений запланированным образовательным результатам.

1.4.1 Примеры тестовых заданий:

Вопрос	Варианты ответов
Что такое модуль вектора в декартовой системе координат?	1. длина вектора 2. угол поворота вектора относительно оси Ox 3. сумма координат вектора 4. сумма модулей координат вектора
Что из нижеперечисленного является логическим высказыванием?	1. «Чему равен определитель матрицы?» 2. «Определитель матрицы равен 0» 3. «Обратная матрица может существовать только у невырожденных матриц» 4. «Модуль матрицы» 5. «Матрица логических высказываний»
Определите по матрице свойства бинарного отношения $M = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$	1) Транзитивность 2) Рефлексивность 3) Симметричность 4) Антирефлексивность 5) Антисимметричность 6) Полнота

Пример типового практического задания:

Создайте функцию, реализующую алгоритм приведения формулы, заданной списком ее значений на всех наборах переменных, к СКНФ. Проверьте работу

функции с помощью библиотечной функции `sympy.logic.boolalg.is_cnf()`.

Данное практическое задание предполагает выполнение следующих этапов:

- 1) Задать формулу списком ее значений на всех наборах переменных.
- 2) Реализовать алгоритм конструирования элементарных дизъюнкций на основе набора переменных.
- 3) Реализовать алгоритм конструирования Совершенной Конъюнктивной Нормальной Формы из элементарных дизъюнкций на основе списка значений данной формулы.
- 4) Проверить работу реализованных функции с помощью библиотечной функции `sympy.logic.boolalg.is_cnf()`.
- 5) Оформить и защитить отчет по практическому заданию.

1.5 Образцы учебно-методических материалов для обучающихся и преподавателей

Для теоретической подготовки используются учебная литература и комплект теоретических материалов, в виде электронных презентаций к лекциям, содержащие основные определения, математический аппарат и иллюстративный материал.

Для подготовки к лабораторным работам используется лабораторный практикум, содержащий описание методов и подходов к выполнению лабораторных заданий.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

1.6 Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE). Возможен формат BYOD.

1.7 Информационное обеспечение реализации рабочей программы

Перечень учебной литературы

- 1) Бугров Я.С. Высшая математика: В 3-х т.: Учеб. для вузов. Т. 1: Дифференциальное и интегральное исчисление: в 2 кн.: Кн. 1 / Я.С. Бугров, С.М. Никольский. — 7-е изд., стер. — М.: Юрайт, 2020. — 253 с.
- 2) Ильин В.А. Линейная алгебра: Учебник для вузов / Ильин В.А., Позняк Э.Г. — 6-е изд., стер. — М.: Физматлит, 2010. — 278 с. — (Курс высшей математики и математической физики. Вып. 4)
- 3) Сборник задач по математике для втузов. В 4-х частях: Учебное пособие для втузов. /Под общ. ред. А.В. Ефимова и А.С. Поспелова. — 5-е изд. испр. — М.: Физматлит, 2009. — Ч. 1.
- 4) Бронштейн И.Н. Справочник по математике для инженеров и учащихся вузов: Учеб. пособие / И.Н. Бронштейн, К.А. Семендяев. — СПб.: Лань, 2010. — 608 с.

Информационные ресурсы

- 1) Интерактивная среда для выполнения Python-кода с возможностью визуализации и документирования. URL: <https://jupyter.org> (дата обращения: 27.07.2025).
- 2) Официальная документация Python 3. URL: <https://docs.python.org/3> (дата обращения: 27.07.2025).

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

2. Рабочая программа модуля

Python для анализа данных

(наименование модуля)

Москва 2025

2.1 Область применения рабочей программы

Рабочая программа модуля «Python для анализа данных» (далее — рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки «Компьютерная лингвистика». Рабочая программа нацелена на формирование у обучающихся знаний и навыков программирования на языке Python для решения прикладных задач.

Программа формирует компетенцию **ID 28 Применяет языки программирования для решения профессиональных задач** на продвинутом уровне: самостоятельно применяет языки программирования и настраиваемые программные инструменты для автоматизации процессов в профессиональной деятельности.

Индикаторы достижения компетенции:

Знать (З2): синтаксис Python и основные конструкции; методы работы с файлами и функциями; принципы объектно-ориентированного программирования; методы обработки ошибок; основы многопоточности и сетевого взаимодействия.

Уметь (У2): писать программы на Python для анализа данных; использовать библиотеки Python для обработки и анализа информации; проектировать и реализовывать классы и модули; отлаживать и тестировать код; применять многопоточность и создавать сетевые приложения.

Программа является обязательной к освоению.

2.2 Структура и краткое содержание рабочей программы

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1	Тема 1. Введение в Python Лекция 1. История программирования, сферы применения, особенности синтаксиса, установка и настройка среды (PyCharm, VS Code, Jupyter Notebook)	2 (лекции) 4 (лабораторные) 2 (СРС)

	Лабораторная работа 1. Настройка окружения, «Hello, world!», работа в интерактивном режиме Самостоятельная работа: изучить официальный туториал Python, подготовить список примеров применения Python в своей области	
2	Тема 2. Ввод и вывод данных Лекция 2. Методы <code>input()</code>, <code>print()</code>, форматирование строк, escape-последовательности Лабораторная работа 2. Программа ввода/вывода с форматированием, создать программу, принимающую имя и возраст и выводящую приветствие Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий	2 (лекции) 4 (лабораторные) 2 (СРС)
3	Тема 3. Работа с файлами Лекция 3. Чтение и запись файлов, режимы, <code>with open</code> Лабораторная работа 3. Программа чтения данных и записи результата в новый файл, прочитать CSV, извлечь нужные столбцы, записать в отдельный файл. Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий	2 (лекции) 4 (лабораторные) 2 (СРС)
4	Тема 4. Работа с функциями Лекция 4. Определение функций, параметры, возвращаемые значения, аннотации типов Лабораторная работа 4. Функции для обработки данных (сумма, среднее, фильтрация), создать библиотеку функций для анализа текста. Самостоятельная работа: изучение учебных материалов и	2 (лекции) 4 (лабораторные) 2 (СРС)

	выполнение текущих домашних работ по темам лабораторных занятий	
5	Тема 5. ООП в Python Лекция 5. Классы, объекты, наследование, полиморфизм, инкапсуляция, магические методы Лабораторная работа 5. Класс для работы с данными (загрузка, фильтрация, сортировка), создать два класса, моделирующие предметную область. Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий	2 (лекции) 4 (лабораторные) 2 (СРС)
6	Тема 6. Анализ ошибок Лекция 6. Исключения, try/except/finally, собственные исключения Лабораторная работа 6. Программа с обработкой ошибок при чтении файла, модифицировать программу для корректной работы при ошибках ввода Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий	2 (лекции) 4 (лабораторные) 2 (СРС)
7	Тема 7. Прикладные задачи Лекция 7. Библиотеки pandas, numpy Лабораторная работа 7. Анализ данных из CSV и генерация отчета, решение задач анализа данных с pandas и matplotlib Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий	2 (лекции) 4 (лабораторные) 2 (СРС)
8	Тема 8. Многопоточность и сетевые приложения Лекция 8. Обзор: threading, concurrent.futures, socket,	2 (лекции) 4 (лабораторные)

	клиент-сервер, простые протоколы Лабораторная работа 8. Добавить многопоточность в существующую программу. Самостоятельная работа: изучение учебных материалов и выполнение текущих домашних работ по темам лабораторных занятий	2 (СРС)
	Промежуточная аттестация в формате тестирования и решения практического задания	2

2.3 Учебно-тематический план рабочей программы

№ п/ п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов			
		аудиторных		самостоятельной работы	
		всего, часов	практические занятия	всего, часов	практические занятия
1	Тема 1. Введение в Python	6	4	2	1
2	Тема 2. Ввод и вывод данных	6	4	2	1
3	Тема 3. Работа с файлами	6	4	2	1
4	Тема 4. Работа с функциями	6	4	2	1
5	Тема 5. ООП в Python	6	4	2	1
6	Тема 6. Анализ ошибок	6	4	2	1
7	Тема 7. Прикладные задачи	6	4	2	1
8	Тема 8. Многопоточность и сетевые приложения	6	4	2	1

	Промежуточная аттестация	2
	Итого	66

2.4 Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение текущего, промежуточного контроля. Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Для текущего и промежуточного контроля создаются фонды оценочных средств (ФОС). ФОС включают в себя педагогические контрольно-измерительные материалы, предназначенные для определения соответствия (или несоответствия) индивидуальных образовательных достижений запланированным образовательным результатам.

2.4.1 Примеры оценочных средств

Примеры тестовых заданий:

Вопрос	Варианты ответов
Какой тип данных в Python изменяемый?	A) tuple B) str C) list D) int
Какой метод удаляет пробелы в начале и конце строки?	A) cut() B) trim() C) strip() D) remove()

Какой модуль используется для многопоточности?	A) multiprocessing B) threading C) asyncio D) parallel
Что такое инкапсуляция в ООП?	A) Запрет наследования B) Объединение данных и методов в одном классе C) Автоматическое создание методов D) Копирование объектов
Как в Python обрабатываются исключения?	A) try/finally B) try/catch C) try/except D) error handling block

Примеры типовых практических заданий:

1. Напишите функцию, которая принимает список чисел и возвращает их среднее значение.
2. Реализуйте класс Employee с методами добавления и отображения информации о сотруднике.
3. Разработайте скрипт, который скачивает данные по API и сохраняет их в файл.
4. Напишите многопоточную программу, которая обрабатывает несколько файлов одновременно.
5. Создайте простое клиент-серверное приложение для обмена сообщениями.

2.5 Учебно-методические материалы для обучающихся и преподавателей

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный

практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

2.6 Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE). Предусмотрен формат BYOD.

2.7 Информационное обеспечение реализации рабочей программы

Список литературы:

1. Сузи, Р. А. Язык программирования Python: учебное пособие / Р. А. Сузи. — 2-е изд. — Москва: ИНТУИТ, 2016. — 350 с. — URL: <https://e.lanbook.com/book/100546> (дата обращения: 19.07.2025). - Режим доступа: для авториз. пользователей МИЭТ.
2. Лутц М. Изучаем Python. — 5-е изд. — СПб.: Питер, 2022. — 1600 с.
3. Седжвик Р., Уэйн К. Алгоритмы на Python. — СПб.: Питер, 2021. — 928 с.
4. Beazley D. Python Cookbook. — 4th ed. — O'Reilly Media, 2023. — 706 p.
5. Downey A. Think Python: How to Think Like a Computer Scientist. — 3rd ed. — Green Tea Press, 2023. — 300 p.

Информационные ресурсы

1. Интерактивная среда для выполнения Python-кода с возможностью визуализации и документирования. URL: <https://jupyter.org> (дата обращения: 27.07.2025).
2. Официальная документация Python 3. URL: <https://docs.python.org/3> (дата обращения: 27.07.2025).

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

3. Рабочая программа модуля

Структурная лингвистика

(наименование модуля)

3.1. Область применения рабочей программы

Рабочая программа модуля «Структурная лингвистика» (далее — рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки «Компьютерная лингвистика». Рабочая программа нацелена на изучение звуковой, морфологической и синтаксической структуры языка, рассматриваются основные фонологические теории, основы морфологии и генеративной грамматики, семантические типы и отношения, а также методы обработки языковых структур с помощью статистических и машинных моделей. Целью программы является развитие у студентов понимания теоретических моделей и формирование навыков применения алгоритмов для анализа реальных лингвистических данных. Рабочая программа нацелена на формирование компетенции **ID 37 Участвует в проектах с применением искусственного интеллекта и машинного обучения под контролем опытных специалистов** на продвинутом уровне: разрабатывает отдельные части проектов по применению искусственного интеллекта и машинного обучения.

По результатам освоения модуля слушатель должен

Знать (З4):

- основные понятия и теории фонетики и фонологии, включая глубинные представления, генеративную фонологию и теорию оптимальности;
- строение морфологической системы языка и методы морфологического анализа;
- принципы генеративной и X-штрих грамматики;
- семантические типы, рестрикторы и кванторы и их роль в структуре предложения;
- основные подходы в обработке естественного языка и вероятностные модели.

Уметь (У4):

- проводить фонетический и морфологический анализ слов и предложений;
- строить деревья зависимостей и применять методы генеративной

грамматики;

- использовать библиотеки для NLP для предобработки текста и построения моделей классификации и регрессии;
- разрабатывать и оценивать простые языковые модели;
- анализировать результаты машинного обучения, корректно интерпретировать метрики.

Программа является обязательной для освоения.

3.2 Структура и краткое содержание рабочей программы

№ п/ п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1	Тема 1. Фонетика и фонология Лекция 1. Фонетика; глубинные представления; генеративная фонология; теория оптимальности. Лабораторная работа 1. Введение в курс; обзор библиотек для НЛП; знакомство с IPA и Unicode; простейшая генерация транскрипции Самостоятельная работа: изучение правил фонетической транскрипции; анализ примеров генеративной фонологии, скрипт, преобразующий русские слова в фонетическую запись.	2 (лекции) 4 (лабораторные) 2 (СРС)
2	Тема 2. Фонетика и морфология Лекция 2. Фонетика; теория оптимальности Лабораторная работа 2. Морфологический анализ; деревья зависимостей, выполнение морфологического анализа слов с использованием существующих анализаторов; построение деревьев зависимостей для простых предложений. Самостоятельная работа: изучение литературы по теории оптимальности.	2 (лекции) 4 (лабораторные) 2 (СРС)

3	Тема 3. Грамматика и семантика Лекция 3. Грамматика; генеративная грамматика Лабораторная работа 3. Семантика; логика и векторные представления, построение синтаксических деревьев; реализация простых моделей векторных представлений и вычисление семантической близости. Самостоятельная работа: изучение литературы по основам генеративной грамматики и понятиям логических форм.	2 (лекции) 4 (лабораторные) 2 (СРС)
4	Тема 4. Синтаксис и предобработка Лекция 4. Грамматика; X-штрих теория Лабораторная работа 4. Предобработка текста; основные этапы машинного обучения, подготовка текстового корпуса: токенизация, нормализация, удаление стоп-слов; исследование влияния предобработки на качество анализа. Самостоятельная работа: изучить литературу по теме основные положения X-штрих теории.	2 (лекции) 4 (лабораторные) 2 (СРС)
5	Тема 5. Семантические типы Лекция 5. Семантика; семантические типы Лабораторная работа 5. Классификация; применение простых классификаторов. Создание датасета и реализация модели классификации для лингвистических задач; анализ метрики точности и полноты. Самостоятельная работа: анализ типов семантических категорий.	2 (лекции) 4 (лабораторные) 2 (СРС)
6	Тема 6. Кванторы и регрессия Лекция 6. Семантика; рестрикторы, кванторы, сферы применения	2 (лекции) 4 (лабораторные) 2 (СРС)

	Лабораторная работа 6. Регрессия; построение регрессионных моделей. применение регрессионной модели для оценки параметров в текстовых данных; сравнение результатов разных моделей. Самостоятельная работа: изучение материалов о кванторах и сферах применения.	
7	Тема 7. Подходы в NLP (natural language processing, НЛП) Лекция 7. НЛП; подходы в НЛП; вероятностные модели Лабораторная работа 7. Языковое моделирование; введение в n-gram модели, реализация простой n-gram языковой модели; анализ частотных распределений и энтропии. Самостоятельная работа: изучение материалов о вероятностных и статистических моделях.	2 (лекции) 4 (лабораторные) 2 (СРС)
8	Тема 8. Разведочный анализ Лекция 8. НЛП; введение в машинное обучение, разведочный анализ и обработку признаков. Лабораторная работа 8. Мини-проект: выбор датасета, выполнение разведочного анализа, предобработка, выбор алгоритма машинного обучения, оценка метрик Самостоятельная работа: подготовка презентации по мини-проекту	2 (лекции) 4 (лабораторные) 2 (СРС)
	Промежуточная аттестация в формате тестирования и решения практического задания	2

3.3 Учебно-тематический план рабочей программы

№ п/ п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов			
		аудиторных		самостоятельной работы	
		всего, часов	практич еские занятия	всего, часов	практическ ие занятия
1	Тема 1. Фонетика и фонология	6	4	2	1
2	Тема 2. Фонетика и морфология	6	4	2	1
3	Тема 3. Грамматика и семантика	6	4	2	1
4	Тема 4. Синтаксис и предобработка	6	4	2	1
5	Тема 5. Семантические типы	6	4	2	1
6	Тема 6. Кванторы и регрессия	6	4	2	1
7	Тема 7. Подходы в NLP (natural language processing, НЛП)	6	4	2	0
8	Тема 8. Разведочный анализ	6	4	2	2
	Промежуточная аттестация	2			
	Итого	66			

3.4 Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение текущего, промежуточного контроля. Текущий контроль проводится преподавателем на основе оценивания

результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Для текущего и промежуточного контроля создаются фонды оценочных средств (ФОС). ФОС включают в себя педагогические контрольно-измерительные материалы, предназначенные для определения соответствия (или несоответствия) индивидуальных образовательных достижений запланированным образовательным результатам.

3.4.1 Примеры оценочных средств

Примеры тестовых заданий:

Вопрос	Варианты ответов
Что является основной единицей фонологического анализа в структурной лингвистике?	а) Морфема б) Фонема с) Аллофон d) Графема
Какой метод используется для выявления минимальных смыслоразличительных единиц языка?	а) Метод трансформаций б) Метод компонентного анализа с) Метод оппозиции d) Метод статистического анализа
Кто считается основателем Пражской лингвистической школы?	а) Ноам Хомский б) Роман Якобсон с) Фердинанд де Соссюр d) Леонард Блумфилд
Что в структурной лингвистике понимается под дистрибуцией?	а) Распределение языковых единиц в тексте б) Соотношение звуков в фонологической системе

	с) Набор морфологических парадигм d) Порядок слов в предложении
Как называется процесс выделения синтагм в потоке речи?	а) Сегментация b) Ассимиляция с) Деривация d) Редукция

Примеры типовых практических заданий:

1. Проанализируйте *приведённый* отрывок текста, выделите фонемы и укажите их оппозиции.
2. Постройте морфологическую схему слова «непрочитанный», обозначив корень, приставку, суффикс и окончание.
3. Произведите синтаксический разбор предложения «Студенты анализируют текст с точки зрения его структуры».
4. Используя метод компонентного анализа, опишите семантические различия между словами «дом» и «здание».
5. Разработайте схему структурных уровней языка, от фонетического до прагматического, и кратко опишите каждый уровень.

3.5 Учебно-методические материалы для обучающихся и преподавателей

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций к лекциям, содержащие основные определения, терминологический аппарат и иллюстративный материал. Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

3.6 Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда

разработки (Jupiter, Notepad++, VS CODE). Предусмотрен формат BYOD.

3.7 Информационное обеспечение реализации рабочей программы

Перечень литературы

1. Керстен Гробок, Катрин Хандке. *Введение в структурную лингвистику*. — М.: Языки славянской культуры, 2010.
2. Н. Хомский. *Аспекты теории синтаксиса*. — М.: Прогресс, 1972.
3. П. Адлер. *Математическая лингвистика*. — М.: Едиториал УРСС, 2005.
4. Daniel Jurafsky, James H. Martin. *Speech and Language Processing*. 3rd ed. — Prentice Hall, 2023.
5. Christopher Manning, Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. — MIT Press, 1999.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

4. Рабочая программа модуля

Функциональные и когнитивные модели в лингвистике

(наименование модуля)

Москва 2025

4.1 Область применения рабочей программы

Рабочая программа модуля «Функциональные и когнитивные модели в лингвистике» (далее — рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки «Компьютерная лингвистика». Программа направлена на получение студентами знаний и умений в области функционального и когнитивного анализа языка, включая изучение принципов функционализма, системно-функциональной лингвистики, фреймовой семантики, когнитивной грамматики и корпусных методов исследования.

Рабочая программа нацелена на получение компетенции **ID 15 Применяет программные алгоритмы обработки данных для инженерно-научных расчетов** на продвинутом уровне: применяет программные алгоритмы обработки данных для инженерно-научных расчетов, самостоятельно подбирает и использует для научных вычислений специализированные библиотеки. Разрабатывает программные алгоритмы с использованием методов математической оптимизации.

По результату освоения дисциплины слушатель должен

Знать (З6):

- основные принципы функционализма (объяснение формы через функцию, типологическая ориентированность, эмпиризм);
- теорию фреймов Ч. Филлмора и её применение в лингвистике;
- основные модели грамматики конструкций (Филлмора, Лакоффа-Гольдберга, Крофта);
- технологии работы с лингвистическими корпусами;
- корпусные методы исследования;
- основные лингвистические ресурсы и базы данных (FrameNet, НКРЯ).

Уметь (У6):

- применять принципы функциональной и когнитивной лингвистики при анализе языкового материала;

- использовать корпусные методы исследования для решения поставленных лингвистических задач;
- работать с лингвистическими ресурсами (FrameNet, НКРЯ) для извлечения и систематизации данных;
- разрабатывать простые аналитические алгоритмы для обработки языковых данных с использованием стандартных библиотек;
- применять модели грамматики конструкций и фреймовой семантики для интерпретации и описания языковых явлений.

Программа является обязательной для освоения.

4.2 Структура и краткое содержание рабочей программы

№ п/ п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1	Тема 1. Основные принципы функционализма Лекция 1. Принципы функционализма: объяснение языковой формы через функции, типологическая ориентированность, эмпиризм, количественные методы, междисциплинарность, иконизм, экономия, мотивация грамматики дискурсом. Проблема грамматикализации. Лабораторная работа 1. Исследование экономии и иконичности в языке на основе корпусных данных. Выявление примеров проявления экономии и иконичность количественными методами в текстах различных стилей, жанров и эпох. Исследование явления экономии: подсчет частотности сокращений и использования клише, сравнение длины слов и предложений в текстах разных стилей, подсчет статистики использования эллипсиса. Исследование	2 (лекции) 4 (лабораторные) 2 (СРС)

	явления иконичности: выявление лексических и морфологических повторов, а также закономерностей их использования в текстах различных стилей; исследование порядка слова и их соответствия последовательности передаваемых событий; изучение явления звукоподобия. Лабораторная работа 2. Исследование принципа грамматикализации (перехода лексических единиц в грамматические) на основе корпусных данных. Частотный анализ: сравнение частотности слова в исходном и грамматикализованном значении. Анализ проявления грамматикализации в диахронии: контекстуальный анализ употребления слова в разных временных срезах. Поиск в корпусе примеров грамматикализации разных типов: фонетическая редукция, дисемантизация, расширение контекстов употребления, фиксация синтаксической позиции. Самостоятельная работа: подготовка отчета по анализу корпусных данных в таблицах и графиках.	
2	Тема 2. Системно-функциональная лингвистика (СФЛ) Лекция 2. Модель языка в СФЛ. Метафункции: понятийная, межличностная, текстуальная. Кардиффская модель функциональной грамматики (М. Халлидей и др.), лексико-грамматика, классы и функции, грамматикализация. Лабораторная работа 1. Анализ текста с точки зрения метафункций СФЛ. Лабораторная работа 2. Применение кардиффской модели для анализа предложений.	2 (лекции) 4 (лабораторные) 2 (СРС)

	Самостоятельная работа: сравнительный анализ текстов с использованием СФЛ.	
3	Тема 3. Функциональная грамматика Лекция 3. Функциональная грамматика на уровне предложения: предложение как сообщение; информационная структура предложения; предложение как диалог; предложение как репрезентация. Функциональная грамматика на уровне дискурса. Лабораторная работа 1. Разбор информационной структуры текста. Лабораторная работа 2. Анализ дискурса с позиций функциональной грамматики. Выявление средств связности в тексте, анализ роли грамматических структур в организации дискурса. Самостоятельная работа: анализ информационной структуры и выявление дискурсивных маркеров в заданном тексте.	2 (лекции) 4 (лабораторные) 4 (СРС)
4	Тема 4. Фреймовая семантика и когнитивные модели Лекция 4. Понятие фрейма. Фреймовая семантика Ч. Филлмора. Структура фрейма. Сцены и скрипты. Фреймы в дискурсе. Фреймовый анализ с применением корпусов. Фреймовый подход к анализу терминологии (П. Фабер и др.). Лабораторная работа 1. Фреймовый анализ терминологии с использованием корпусов. Лабораторная работа 2. Построение схемы предметной онтологии на основе фреймов. Самостоятельная работа: разработка плана создания фреймовой онтологии с помощью средств компьютерной лингвистики.	2 (лекции) 4 (лабораторные) 4 (СРС)

5	Тема 5. Идеализированные когнитивные модели (ИКМ) Лекция 5. Кластерные и метонимические ИКМ. Лингвокультурная специфика ИКМ. Когнитивный подход к лексической семантике: полисемия, гиперо- и гипонимия, антонимия. Проект FrameNet и аналогичные проекты. Опыт применения фреймового анализа для построения предметных онтологий. Лабораторная работа 1. Сбор данных для построения формальной онтологии. Лабораторная работа 2. Создание скрипта для разработки формальной онтологии. Самостоятельная работа: разработка формальной онтологии с помощью средств компьютерной лингвистики.	2 (лекции) 4 (лабораторные) 2 (СРС)
6	Тема 6. Когнитивная грамматика Лекция 6. Субъективистский подход к значению. Профиль - база, траектор - ориентир, когнитивная область. Когнитивные точки отсчета. Метонимия. Активная зона. Когнитивный подход к исследованию грамматических явлений: время, вид, залог, модальность с точки зрения когнитивной лингвистики. Лабораторная работа 1. Анализ концептов с точки зрения понятий когнитивной грамматики: определение терминов «профиль», «база», «траектор», «ориентир» с использованием языковых примеров и схем. Лабораторная работа 2. Анализ грамматических категорий (время, вид, залог) в заданном тексте с когнитивной точки зрения. Самостоятельная работа: анализ явления метонимии на	2 (лекции) 4 (лабораторные) 2 (СРС)

	материале корпусов.	
7	Тема 7. Грамматика конструкций Лекция 7. Модели Филлмора, Лакоффа-Гольдберга, У. Крофта. Коллострукционный анализ. Грамматика конструкции и понятие коллострукции. Коллострукционный анализ с применением корпусных данных. Речевой опыт как основа формирования и функционирования грамматических категорий. Лабораторная работа 1. Применение модели Филлмора для анализа конструкций в заданном тексте. Лабораторная работа 2. Корпусный анализ коллокаций и коллострукций. Самостоятельная работа: применение модели Филлмора для анализа конструкций в заданном тексте.	2 (лекции) 4 (лабораторные) 2 (СРС)
8	Тема 8. Корпусные методы в когнитивной лингвистике Лекция 8. Корпусный подход в когнитивной лингвистике. Анализ статистических данных в когнитивной лингвистике. Эффект частотности. Корпусный анализ коллокаций в когнитивной лингвистике. Коллексемный анализ. Инструменты для корпусных исследований. Лабораторная работа 1. Использование корпусных инструментов (AntConc, Sketch Engine и др.) для анализа данных. Сравнение существующих инструментов для анализа корпусов с возможностями инструментов из области анализа данных. Лабораторная работа 2. Сравнительный анализ частотности языковых явлений в разных корпусах: выбор явления и составление сравнительно-сопоставительных	2 (лекции) 2 (лабораторные) 2 (СРС)

	таблиц и графиков. Самостоятельная работа: подготовка исследования на основе корпусных данных. Разработка инструментария для проведения корпусного исследования (анализ существующих инструментов или разработка новых инструментов).	
9	Промежуточная аттестация в формате тестирования и решения практического задания	2

4.3 Учебно-тематический план рабочей программы

№ п/ п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов			
		аудиторных		самостоятельной работы	
		всего, часов	практические занятия	всего, часов	практические занятия
1	Тема 1. Основные принципы функционализма	6	4	2	1
2	Тема 2. Системно-функциональная лингвистика (СФЛ)	6	4	2	1
3	Тема 3. Функциональная грамматика	6	4	4	3
4	Тема 4. Фреймовая семантика и когнитивные модели	6	4	4	3
5	Тема 5. Идеализированные когнитивные модели (ИКМ)	6	4	2	1
6	Тема 6. Когнитивная грамматика	6	4	2	1

7	Тема 7. Грамматика конструкций	6	4	2	1
8	Тема 8. Корпусные методы в когнитивной лингвистике	6	4	2	1
	Промежуточная аттестация	2			
	Итого	70			

4.4 Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение текущего, промежуточного контроля. Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в формате тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Для текущего и промежуточного контроля создаются фонды оценочных средств (ФОС). ФОС включают в себя педагогические контрольно-измерительные материалы, предназначенные для определения соответствия (или несоответствия) индивидуальных образовательных достижений запланированным образовательным результатам.

4.4.1 Примеры оценочных средств

Примеры тестовых заданий:

Вопрос	Варианты ответов
Какой из перечисленных принципов НЕ относится к функционализму?	а) Объяснение языковой формы через её функции б) Генеративистский подход (акцент на врождённых структурах) в) Принцип экономии

	г) Мотивация грамматики дискурсом
Какое из утверждений верно характеризует проблему грамматикализации?	а) Это процесс превращения лексических единиц в грамматические маркеры. б) Это изучение только синхронных изменений в языке. в) Это анализ исключительно фонетических закономерностей. г) Это теория, отрицающая влияние дискурса на грамматику.
Какие три метафункции языка выделяются в СФЛ?	а) Фонетическая, морфологическая, синтаксическая б) Понятийная, межличностная, текстуальная в) Денотативная, коннотативная, экспрессивная г) Референциальная, эмоциональная, поэтическая
Какой элемент НЕ входит в кардиффскую модель функциональной грамматики?	а) Лексико-грамматика б) Грамматикализация в) Генеративные правила трансформации г) Классы и функции
Что такое фрейм в семантике Ч. Филлмора?	а) Морфологическая схема словообразования б) Когнитивная структура, организующая знания о типичной ситуации в) Синтаксическое дерево зависимостей

	г) Статистическая модель распределения слов в корпусе
Какой проект использует фреймовый подход для анализа лексики?	а) WordNet б) FrameNet в) BERT г) GPT
Что означает «профиль-база» в когнитивной грамматике?	а) Соотношение активной зоны значения и его когнитивной основы б) Частотность употребления слова в корпусе в) Морфологическую парадигму слова г) Синтаксическую функцию слова в предложении
Какая модель НЕ относится к грамматике конструкций?	а) Модель Филлмора б) Модель Хомского в) Модель Лакоффа-Гольдберга г) Модель У. Крофта
Что изучает коллокационный анализ?	а) Распределение фонем в тексте б) Закономерности совместной встречаемости слов и грамматических структур в) Частотность отдельных морфем г) Диалектные различия в произношении
Какой инструмент НЕ используется для корпусного анализа?	а) AntConc б) Sketch Engine в) ELAN г) Word2Vec

Примеры типовых практических заданий:

1) Произведите корпусной анализ с целью выявления примеров проявления экономии и иконичности в языке:

а) Найдите примеры проявления экономии в текстах XIX и XX веков: посчитайте частотность сокращений и использования клише, сравните длины слов и предложений в текстах, посчитайте статистики использования эллипсиса.

б) Найдите примеры проявления иконичности в текстах XIX и XX веков: найдите образцы использования лексических и морфологических повторов, опишите закономерности их использования; проанализируйте соответствие использования порядка слова при передаче последовательности передаваемых событий; найдите примеры использования звукоподобия в текстах.

2) Сравните результаты в текстах XIX и XX века, опишите выявленные закономерности и ключевые различия в проанализированных текстах.

4.5 Учебно-методические материалы для обучающихся и преподавателей

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций, содержащие основные определения, терминологический аппарат и иллюстративный материал.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

4.6 Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE). Предусмотрен формат BYOD.

4.7 Информационное обеспечение реализации рабочей программы

Перечень учебной литературы

1. Halliday M.A.K. An Introduction to Functional Grammar. – Oxford University Press Inc, 2004. URL: https://www.uel.br/projetos/ppcat/pages/arquivos/RECURSOS/2004_HALLIDAY_MATTHIESSEN_An_Introduction_to_Functional_Grammar.pdf (дата обращения 04.08.2025)
2. Fillmore C. J. Frame Semantics. – Fillmore Department of Linguistics University of California Berkeley, 1976. URL: https://brencon.com/Fillmore%201982_2up.pdf (дата обращения 04.08.2025)
3. Лакофф Д. Женщины, огонь и опасные вещи: Что категории языка говорят нам о мышлении / Пер. с англ. И. Б. Шатуновского. — М.: Языки славянской культуры, 2004. — 792 с. — (Язык. Семиотика. Культура). URL: https://psv4.userapi.com/s/v1/d/GqJZgMIGfm-5ba0Vqe-42RUHqS1tTXLxGPGTMUsIkp7HlImpjMXzymfD8ubXAqPIU03SJSbxAkz9hd_PZmvpOOZhFmUDv97uH0Dfti7oPnRylRd/2004_Lakoff_Dzh_Zhenschiny_ogon_i_opasnye_veschi_Chto_kategorii_yazyka_govoryat_nam_o_myshlenii.pdf (дата обращения 04.08.2025)
4. Кибрик А.А. Когнитивные исследования по дискурсу. — М.: 1994. — № 5. — С. 126-139. URL: <https://vja.ruslang.ru/ru/archive/1994-5/126-139> (дата обращения 04.08.2025)

Информационные ресурсы

Cognitive linguistics. URL: <https://www.cognitivelinguistics.org/en/about-cognitive-linguistics> (дата обращения: 04.08.2025).

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

5. Рабочая программа модуля

SQL и NoSQL базы данных

(наименование модуля)

Москва 2025

5.1 Область применения рабочей программы

Рабочая программа модуля «SQL и NoSQL базы данных» (далее — рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки «Компьютерная лингвистика». Рабочая программа нацелена на формирование у обучающихся знаний и навыков проектирования и использования SQL и NoSQL баз данных, построения запросов, работы с данными и интеграции СУБД с клиентскими приложениями.

Изучение модуля является обязательным и ставит следующие задачи:

- изучить основные модели данных и архитектуру СУБД;
- освоить основы реляционной алгебры;
- овладеть синтаксисом и конструкциями языка SQL;
- научиться составлять сложные SQL-запросы;
- познакомиться с подходами к работе с NoSQL базами данных;
- отработать навыки интеграции СУБД с клиентскими приложениями.

Программа формирует компетенцию **ID 32 Использует СУБД при разработке ПО** на базовом уровне: **знает основы баз данных, знаком с нормализацией, ACID, транзакциями, может написать простые выборки («select'ы»).** Участвует в проектах по созданию ПО с использованием СУБД под контролем опытных специалистов.

Индикаторы достижения компетенции:

Знать (З7): классификацию и архитектуру СУБД; основные модели данных и модель ANSI/SPARC; основы реляционной алгебры; синтаксис SQL и его основные конструкции; принципы работы NoSQL баз данных.

Уметь (У7): проектировать и создавать структуры баз данных; формировать SQL-запросы различной сложности; выполнять операции выборки, обновления и удаления данных; интегрировать СУБД с клиентскими приложениями; использовать NoSQL базы данных для решения прикладных задач.

Программа является обязательной для освоения.

5.2 Структура и краткое содержание рабочей программы

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1	Тема 1. Модели данных. Файловые системы и базы данных. Лекция 1. Понятие модели данных, виды (иерархическая, сетевая, реляционная, документоориентированная); отличие файловых систем от СУБД, преимущества использования СУБД. Лабораторная работа 1. Создание и настройка БД. Установка и настройка PostgreSQL/MySQL; подключение к серверу; создание базы данных, схем и таблиц; определение типов данных; наполнение тестовыми данными; выполнение базовых CRUD-операций. Самостоятельная работа: изучение учебных материалов по темам лабораторных занятий	2 (лекции) 4 (лабораторные) 4 (СРС)
2	Тема 2. Модель ANSI/SPARC. СУБД. Лекция 2. Уровни модели: внешний, концептуальный, внутренний; архитектура и классификация СУБД; примеры популярных систем (PostgreSQL, MySQL, MongoDB). Лабораторная работа 2. Реляционная алгебра и базовые запросы SQL. Выполнение операций выборки, соединения, фильтрации, группировки; использование WHERE, ORDER BY, GROUP BY, HAVING; работа с агрегатными функциями. Самостоятельная работа: изучение учебных материалов по темам лабораторных занятий	2 (лекции) 4 (лабораторные) 4 (СРС)
3	Тема 3. Основы реляционной алгебры. SQL: синтаксис	2 (лекции)

	и основные операторы. Лекция 3. Операции выборки, проекции, соединения, объединения; структура SQL-запроса; операторы SELECT, INSERT, UPDATE, DELETE; агрегатные функции и группировка данных. Лабораторная работа 3. Взаимодействие СУБД с клиентскими приложениями. Подключение БД из Python/Java; выполнение SQL-запросов из приложения; обработка результатов и вывод отчёта. Самостоятельная работа: изучение учебных материалов по темам лабораторных занятий	4 (лабораторные) 4 (СРС)
4	Тема 4. Сложные запросы и NoSQL. Лекция 4. Подзапросы, CTE, индексы и оптимизация запросов; обзор NoSQL (MongoDB), особенности документоориентированных БД, базовые CRUD-операции в MongoDB. Лабораторная работа 4. Сложные запросы SQL. Подзапросы, CTE, оконные функции, использование индексов; анализ планов выполнения запросов (EXPLAIN). Лабораторная работа 5. Работа с NoSQL БД. Создание коллекций в MongoDB, вставка, выборка, обновление, удаление документов; фильтрация и сортировка данных; вложенные документы. Самостоятельная работа: изучение учебных материалов по темам лабораторных занятий	2 (лекции) 8 (лабораторные) 8 (СРС)
5	Промежуточная аттестация в формате тестирования и решения практического задания	2

5.3 Учебно-тематический план рабочей программы

№ п/ п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов			
		аудиторных		самостоятельной работы	
		всего, часов	практич еские занятия	всего, часов	практич еские занятия
1	Тема 1. Модели данных. Файловые системы и базы данных.	6	4	4	2
2	Тема 2. Модель ANSI/SPARC. СУБД.	6	4	4	2
3	Тема 3. Основы реляционной алгебры. SQL: синтаксис и основные операторы.	6	4	4	2
4	Тема 4. Сложные запросы и NoSQL.	10	8	8	4
	Промежуточная аттестация	2			
	Итого	50			

5.4 Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение текущего, промежуточного контроля. Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Для текущего и промежуточного контроля создаются фонды оценочных

средств (ФОС). ФОС включают в себя педагогические контрольно-измерительные материалы, предназначенные для определения соответствия (или несоответствия) индивидуальных образовательных достижений запланированным образовательным результатам.

5.4.1 Примеры оценочных средств

Примеры тестовых заданий:

Вопрос	Варианты ответов
Какая операция реляционной алгебры используется для выборки определённых столбцов?	A) Join B) Select C) Project D) Union
Какой оператор SQL используется для сортировки результата?	A) SORT BY B) ORDER BY C) GROUP BY D) ARRANGE
Что означает уровень «внешний» в модели ANSI/SPARC?	A) Физическая реализация данных B) Представление данных для пользователя C) Логическая структура базы D) Оптимизация запросов
Какая СУБД является документоориентированной?	A) MySQL B) PostgreSQL C) MongoDB D) Oracle
Какой оператор используется для объединения результатов двух запросов?	A) UNION B) JOIN C) INTERSECT D) MERGE

Примеры практических заданий

- 1) Спроектируйте схему БД для онлайн-магазина, определите ключи.
- 2) Составьте SQL-запрос для получения списка клиентов, оформивших заказы в прошлом месяце.
- 3) Оптимизируйте запрос с использованием индекса.
- 4) Создайте коллекцию MongoDB и добавьте в неё 5 документов.
- 5) Напишите код на Python для подключения к СУБД и выполнения запроса.

5.5 Учебно-методические материалы для обучающихся и преподавателей

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций, содержащие основные определения, терминологический аппарат и иллюстративный материал.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

5.6 Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE). Предусмотрен формат BYOD.

5.7 Информационное обеспечение реализации рабочей программы

Перечень учебной литературы

- 1) Кузнецов, С. Е. Язык SQL. Полное руководство / С. Е. Кузнецов. — 4-е изд., испр. и доп. — Санкт-Петербург: Питер, 2023. — 720 с. — (Серия «Программист профессионал»).
- 2) Миклис, С. В. PostgreSQL 15: руководство администратора и разработчика / С. В. Миклис. — Москва: ДМК Пресс, 2023. — 496 с.

- 3) Чекмарев, А. В. NoSQL для разработчиков: MongoDB, Redis, Cassandra / А. В. Чекмарев. — Москва: БХВ-Петербург, 2022. — 384 с.
- 4) Elmasri, R. Fundamentals of Database Systems / R. Elmasri, S. B. Navathe. — 8th ed. — Boston: Pearson, 2023. — 1200 p.
- 5) Coronel, C. Database Systems: Design, Implementation, and Management / C. Coronel, S. Morris. — 14th ed. — Boston: Cengage, 2022. — 880 p.
- 6) Date, C. J. Database Design and Relational Theory: Normal Forms and All That Jazz / C. J. Date. — 3rd ed. — Beijing: O'Reilly Media, 2021. — 600 p.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

6. Рабочая программа модуля

Автоматическая обработка текста

(наименование модуля)

6.1 Область применения рабочей программы

Рабочая программа модуля «**Автоматическая обработка текста**» (далее — рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки «Компьютерная лингвистика». Рабочая программа нацелена на формирование компетенции **ID 15 Применяет программные алгоритмы обработки данных для инженерно-научных расчётов** в области обработки текстовых данных и построения простых моделей анализа и классификации текста на продвинутом уровне: применяет программные алгоритмы обработки данных для инженерно-научных расчетов, самостоятельно подбирает и использует для научных вычислений специализированные библиотеки. Разрабатывает программные алгоритмы с использованием методов математической оптимизации.

Для освоения компетенции требуется:

Знать (З5): основные этапы обработки текстов: очистка, токенизация, нормализация, лемматизация, морфология, синтаксис, семантика; методы распознавания именованных сущностей и оценки качества; методы морфологического анализа и POS-тегирования, различия между стеммингом и лемматизацией; основы конечных автоматов (FSA/FST) для морфологических задач; принципы работы с API и сбор данных из открытых источников; контекстно-свободные и зависимые грамматики, алгоритмы CYK и Earley, методы извлечения коллокаций; тематическое моделирование (LDA) и меры сходства текстов; методы классификации, суммаризации и анализа тональности текстов; особенности обработки многоязычных текстов и алгоритмы извлечения ключевых слов; основы анализа текстов социальных медиа и меры сходства текстов.

Уметь (У5): выполнять предобработку текста; реализовывать методы rule-based и статистического NER, POS-тегирования и стемминга/лемматизации на Python; строить конечные автоматы для морфологических задач; получать данные через API и web-скрейпинг, соблюдать этические нормы; реализовывать простейшие парсеры на базе CFG и алгоритмы переходного-ориентированного

анализа; применять методы тематического моделирования и вычислять меры сходства; строить классификаторы и проводить анализ тональности; реализовывать экстрактивные алгоритмы суммаризации и простые QA-системы; работать с многоязычными данными и извлекать ключевые слова; выполнять сетевой анализ текстов и оценивать сходство текстов.

Программа является обязательной для освоения.

6.2 Структура и краткое содержание рабочей программы

№ п/ п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1	Тема 1. Введение в обработку текста и естественного языка Лекция 1. Определение и области применения NLP (анализ текста, классификация, анализ комментариев). Текст как данные: структурированные и неструктурированные данные, кодировки (ASCII, Unicode). Этапы обработки текста: предобработка, токенизация, нормализация, морфология, синтаксис, семантика. Предобработка текста: очистка (удаление шума, URL, пунктуации), нормализация регистра, удаление стоп-слов. Введение в анализ текстов: комментарии, большие тексты, сетевой анализ. Python для обработки текста: библиотеки re, string, collections, numpy. Лабораторная работа 1. Чтение/запись текстовых файлов в Python (open, encode, decode). Предобработка: удаление URL, email, чисел, специальных символов с помощью регулярных выражений (re). Токенизация: разделение текста на слова и предложения (правила). Удаление стоп-слов: создание собственного списка стоп-	2 (лекции) 2 (лабораторные) 2 (СРС)

	слов. Самостоятельная работа: изучение учебных материалов по темам лекционных и лабораторных занятий.	
2	Тема 2. Распознавание именованных сущностей и метрики оценки Лекция 2. Распознавание именованных сущностей (NER): правило-based и статистические подходы. Метрики оценки: точность, полнота, F1-мера, матрицы ошибок. Применение NER в анализе комментариев и больших текстов. Корпусная лингвистика: создание и анализ корпусов Лабораторная работа 2. Распознавание именованных сущностей (NER). Содержание: Правило-based NER: использование регулярных выражений и словарей для сущностей (люди, организации). Статистический подход: применение HMM для NER. Самостоятельная работа: изучение учебных материалов по темам лекционных и лабораторных занятий.	2 (лекции) 2 (лабораторные) 2 (СРС)
3	Тема 3. Морфологический анализ и неглубокие векторные представления Лекция 3. Морфология в NLP: флексия, деривация, композиты. Лемматизация vs. стемминг: словарные подходы. Неглубокие векторные представления: «мешок слов» (bag-of-words), TF-IDF, их применение в анализе текстов. Конечные автоматы (FSA) и трансдюсеры (FST) для морфологического анализа. Введение в анализ больших текстов: масштабируемость, обработка объемных данных.	2 (лекции) 4 (лабораторные) 4 (СРС)

	Лабораторная работа 3. Анализ комментариев: особенности коротких текстов (шум, смайлики, неформальность). Продвинутая токенизация: обработка сокращений, смайлов, хэштегов. Стемминг: реализация правило-based стеммера (например, для удаления -ing, -ed). Лемматизация. Лабораторная работа 4. Реализация конечного автомата в Python для морфологических задач (например, распознавание глагольных форм). Правило-based POS-тегирование: создание теггера с использованием словарей и регулярных выражений. Самостоятельная работа: изучение учебных материалов по темам лекционных и лабораторных занятий.	
4	Тема 4. Работа с API и статистические модели Лекция 4. Работа с API: получение текстовых данных через API (например, новостные или социальные медиа API). Лабораторная работа 5. Основы веб-скрейпинга: использование библиотек requests и BeautifulSoup для сбора текстов. Парсинг HTML: извлечение текстов из веб-страниц (например, комментариев, статей). Этические аспекты скрейпинга: ограничения, robots.txt. Понимание структуры веб-данных. Лабораторная работа 6. Получение текстовых данных через API (например, используя requests для публичного API). Реализация HMM в Python для POS-тегирования или NER. Лабораторная работа 7. Классификация текста: бинарные и многоклассовые задачи (например, анализ тональности комментариев). Извлечение признаков:	2 (лекции) 8 (лабораторные) 8 (СРС)

	«мешок слов», TF-IDF, реализация в Python. Лабораторная работа 8. Реализация базового парсера для контекстно-свободных грамматик (например, СΥК). Извлечение коллокаций: точечная взаимная информация (PMI). Самостоятельная работа: изучение учебных материалов по темам лекционных и лабораторных занятий.	
5	Тема 5. Синтаксический анализ и грамматика зависимостей Лекция 5. Синтаксический анализ: контекстно-свободные грамматики (CFG), грамматика зависимостей. Алгоритмы парсинга: СΥК, Earley для CFG; переходно-ориентированный парсинг зависимостей. Грамматика зависимостей: представление структуры предложения в виде графов. Введение в сетевой анализ: графы для анализа текстовых связей. Алгоритмы парсинга: СΥК, Earley для CFG; переходно-ориентированный парсинг зависимостей. Грамматика зависимостей: представление структуры предложения в виде графов. Введение в сетевой анализ: графы для анализа текстовых связей. Лабораторная работа 9. Реализация упрощенной версии LDA (например, на основе term co-occurrence). Частотный анализ: слова, фразы, конкордансы в большом корпусе. Лабораторная работа 10. Экстрактивная суммаризация: выбор ключевых предложений на основе весов. Алгоритмы ранжирования предложений: подсчет частот слов, позиционные метрики. Самостоятельная работа: изучение учебных материалов по темам лекционных и лабораторных	2 (лекции) 4 (лабораторные) 4 (СРС)

	занятий.	
6	Тема 6. Тематическое моделирование и анализ больших текстов Лекция 6. Тематическое моделирование: основы Latent Dirichlet Allocation (LDA), применение в анализе текстов. Анализ больших текстов: масштабируемость, работа с большими корпусами. Меры сходства текста: косинусное сходство, коэффициент Жаккара. Неглубокие векторные представления: TF-IDF для больших текстов. Лабораторная работа 11. Анализ тональности: классификация комментариев на позитивные, негативные, нейтральные. Использование словарей тональности и статистических методов. Лабораторная работа 12. Особенности обработки многоязычных текстов: кодировки, токенизация, морфология. Реализация токенизатора для текста на двух языках (например, английский и русский). Лабораторная работа 13. Извлечение ключевых слов: методы на основе TF-IDF, RAKE (Rapid Automatic Keyword Extraction). Реализация алгоритма RAKE для извлечения ключевых фраз. Самостоятельная работа: изучение учебных материалов по темам лекционных и лабораторных занятий.	2 (лекции) 6 (лабораторные) 6 (СРС)
7	Тема 7. Суммаризация, QA-системы и графы знаний Лекция 7. Суммаризация текстов: экстрактивные и статистические методы. QA-системы: правило-based подходы к вопросно-ответным системам. Графы знаний: структура, создание, применение в NLP (например, для анализа связей в текстах). Перспективы NLP: гибридные	2 (лекции) 2 (лабораторные) 2 (СРС)

	подходы, мультязычная обработка. Лабораторная работа 14. Особенности текстов социальных медиа: неформальность, хэштеги, упоминания. Анализ связей в текстах: построение графа упоминаний или хэштегов. Самостоятельная работа: изучение учебных материалов по темам лекционных и лабораторных занятий.	
9	Промежуточная аттестация в формате тестирования и решения практического задания	2

6.3 Учебно-тематический план рабочей программы

№ п/ п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов			
		аудиторных		самостоятельной работы	
		всего, часов	практические занятия	всего, часов	практические занятия
1	Тема 1. Введение в обработку текста и естественного языка	4	2	2	1
2	Тема 2. Распознавание именованных сущностей и метрики оценки	4	2	2	1
3	Тема 3. Морфологический анализ и неглубокие векторные представления	6	4	4	2
4	Тема 4. Работа с API и статистические модели	10	8	8	4

5	Тема 5. Синтаксический анализ и грамматика зависимостей	6	4	4	2
6	Тема 6. Тематическое моделирование и анализ больших текстов	8	6	6	3
7	Тема 7. Суммаризация, QA-системы и графы знаний	4	2	2	1
	Промежуточная аттестация	2			
	Итого	72			

6.4 Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение текущего, промежуточного контроля. Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме тестирования. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Для текущего и промежуточного контроля создаются фонды оценочных средств (ФОС). ФОС включают в себя педагогические контрольно-измерительные материалы, предназначенные для определения соответствия (или несоответствия) индивидуальных образовательных достижений запланированным образовательным результатам.

6.4.1 Примеры оценочных средств

Примеры тестовых заданий:

Вопрос	Варианты ответов
Что представляет собой токенизация в автоматической обработке текста?	a) Преобразование текста в аудиофайл b) Разделение текста на предложения и слова c) Исправление орфографических ошибок d) Замена слов синонимами
Какой алгоритм чаще всего используется для стемминга в английском языке?	a) Porter Stemmer b) Levenshtein Distance c) K-means clustering d) LDA
Что такое стоп-слова?	a) Слова с наибольшей частотой в тексте b) Слова, которые невозможно токенизировать c) Специальные слова для кодирования текста d) Технические термины
Как называется процесс автоматического определения части речи слова в тексте?	a) POS-tagging b) Lemmatization c) Parsing d) Tokenization
Что делает алгоритм TF-IDF?	a) Строит синтаксическое дерево b) Определяет частоту слова в документе с учётом важности в корпусе

	с) Объединяет слова в биграммы d) Переводит текст на другой язык
--	---

Примеры практических заданий:

- 1) Выполните токенизацию предложенного текста, удалите стоп-слова и представьте результат в виде списка токенов.
- 2) Постройте частотный словарь для текста и выделите 10 наиболее часто встречающихся слов.
- 3) Выполните лемматизацию текста с использованием выбранной библиотеки Python (например, spaCy или pymorphy2).
- 4) Рассчитайте TF-IDF для набора предложений и определите, какие слова являются наиболее информативными.
- 5) Разработайте простой алгоритм для выделения именованных сущностей (Named Entity Recognition) в предложенном тексте.

6.5 Учебно-методические материалы для обучающихся и преподавателей

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций, содержащие основные определения, терминологический аппарат и иллюстративный материал.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

6.7 Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE). Предусмотрен формат BYOD.

6.8 Информационное обеспечение реализации рабочей программы

Перечень литературы

- 1) Маннинг, К., Шютце, Х. Основы статистической обработки естественного языка: учебное пособие / пер. с англ. — Москва: ДМК Пресс, 2021. — 640 с.
- 2) Jurafsky, D., Martin, J. H. Speech and Language Processing. — 3rd ed. — London: Pearson, 2023. — 1024 p.
- 3) Васильева, А. Н., Писаревский, С. А. Введение в автоматическую обработку текста: учебное пособие. — Санкт-Петербург: Лань, 2020. — 224 с.
- 4) Bird, S., Klein, E., Loper, E. Natural Language Processing with Python. — 2nd ed. — Beijing: O'Reilly Media, 2021. — 504 p.
- 5) Ключев, Е. В., Гаврилова, Т. А. Технологии обработки естественного языка: монография. — Москва: Наука, 2022. — 312 с.
- 6) Aggarwal, C. C., Zhai, C. Mining Text Data. — New York: Springer, 2019. — 524 p.
- 7) Переверзев, М. А. Компьютерная лингвистика и обработка текстовой информации: учебное пособие. — Москва: Юрайт, 2023. — 310 с.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

7. Рабочая программа модуля

Нейронные сети в обработке естественного языка

(наименование модуля)

Москва 2025

7.1 Область применения рабочей программы

Рабочая программа модуля «Нейронные сети в обработке естественного языка» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Компьютерная лингвистика» и направлена на получение студентами знаний и умений в области обработки естественного языка с использованием современных нейронных сетей, включая большие языковые модели, методы их сжатия, интеграции и развертывания.

Рабочая программа направлена на освоение компетенции **ID 37 Участвует в проектах с применением искусственного интеллекта и машинного обучения под контролем опытных специалистов** на продвинутом уровне: разрабатывает отдельные части проектов по применению искусственного интеллекта и машинного обучения.

Для освоения компетенции требуется:

Знать (ЗЗ):

- современные методы обработки естественного языка;
- алгоритмы построения векторных представлений текста;
- методы предобработки текстовых данных;
- принципы работы нейронных сетей для обработки естественного языка;
- основы работы с большими языковыми моделями .

Уметь (УЗ):

- применять библиотеки для обработки текстов и работы с нейронными сетями;
- разрабатывать и оптимизировать пайплайны для решения задач в сфере обработки естественного языка;
- работать с API больших языковых моделей;
- разрабатывать программные модули для обработки естественного языка.

Программа является обязательной для освоения.

7.2 Структура и краткое содержание рабочей программы

№ п/ п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1	Тема 1. Введение в обработку естественного языка Лекция 1. История развития сферы обработки естественного языка. Основные задачи области обработки естественного языка. Сферы применения инструментов обработки естественного языка. Лабораторная работа 1. Терминологический аппарат инженера в области обработки естественного языка. Составление глоссария. Токенизация. Стемминг. Лемматизация. Стоп-слова. N-граммы. Корпус. Датасет. Парсинг. Разметка. Препроцессинг. Лабораторная работа 2. Освоение работы с основными библиотеками и программными средствами для лингвистического анализа текста (Mystem, Py morphology, NLTK, SpaCy, Natasha). Самостоятельная работа: произведение сравнительного анализа возможностей, преимуществ и недостатков освоенных библиотек для лингвистического анализа. Подготовка отчета о сферах применения изученных инструментов.	2 (лекции) 4 (лабораторные) 2 (СРС)
2	Тема 2. Векторные представления текста Лекция 2. Необходимость построения векторного представления текста в компьютерной обработке естественного языка. Оценка семантической близости слов. Базовые методы построения векторных представлений (Bag-of-Words, One-Hot Encoding, TF-IDF). Статистические и контекстные методы построения	2 (лекции) 4 (лабораторные) 2 (СРС)

	векторных представлений. Статистические модели (Word2Vec, FastText, GloVe). Контекстные эмбединги (ELMo, BERT). Лабораторная работа 1. Построение векторных представлений для заданного текста с помощью статических моделей. Настройка параметров. Вычисление семантически близких слов на основе векторного сходства. Применение методов снижения размерности матриц для визуализации результатов. Лабораторная работа 2. Построение векторных представлений для заданного текста с помощью контекстных моделей. Настройка параметров, выявление семантически близких слов, визуализация результатов. Самостоятельная работа: сравнение результатов работы статистических и контекстных моделей. Анализ перспектив применения изученных инструментов в разработке с опорой на академические и научно-популярные источники.	
3	Тема 3. Сверточные нейронные сети в обработке естественного языка Лекция 3. Классификация как одна из базовых задач обработки текста. Анализ тональности текста, определение авторства, фильтрация спама, морфологический анализ текста. Классические подходы к решению задачи классификации. Введение в работу сверточных нейросетей. Лабораторная работа 1. Разбор работы сверточных нейросетей на примере обработки изображений. Лабораторная работа 2. Реализация алгоритма анализа тональности текстов отзывов на товары с помощью	2 (лекции) 4 (лабораторные) 2 (СРС)

	сверточных нейросетей. Метрики оценки качества классификации текстов. Самостоятельная работа: разработка системы для анализа тональности текстов отзывов на товары; подготовка данных, настройка модели с помощью библиотек PyTorch или TensorFlow, оценка производительности модели; создание отчета о проделанной работе, анализ ошибок классификатора, выявление преимуществ и недостатков использования сверточных нейросетей в разработке.	
4	Тема 4. Рекуррентные нейронные сети в обработке естественного языка Лекция 4. Моделирование языка и генерация текста: сферы применения и вызовы. Основы теории вероятностей для построения моделей для предсказания следующего слова. Рекуррентная нейросеть. Проблема взрывающегося и исчезающего градиента. Проблема «бутылочного горлышка». Долгая краткосрочная память. Управляемый рекуррентный блок. Лабораторная работа 1. N-граммные языковые модели. Лабораторная работа 2. Применение рекуррентных нейросетей в машинном переводе и в разработке диалоговых систем. Самостоятельная работа: обучение рекуррентной нейросети с помощью библиотек PyTorch или TensorFlow для генерации коротких текстов; настройка гиперпараметров модели; оценка качества генерации текста на основе метрики перплексии; анализ влияния качества обучающей выборки на результат генерации.	2 (лекции) 4 (лабораторные) 2 (СРС)

5	Тема 5. Механизм внимания и архитектура «Трансформер» Лекция 5. Архитектура Sequence-to-Sequence. Энкодеры и декодеры в машинном переводе. Механизм внимания: история возникновения, основные принципы работы, применение в нейросетевом машинном переводе. Архитектура «Трансформер» для машинного перевода. Основные компоненты архитектуры: внутреннее внимание, многоголовое внимание, блоки кодера и декодера, позиционное кодирование. Лабораторная работа 1. Исследование основных компонентов архитектуры «Трансформер» и используемых в ней вычислений. Запуск и оценка обученной модели; сравнение результатов работы модели со сверточными и рекуррентными нейросетями на задачах классификации, реферирования текста и машинного перевода. Лабораторная работа 2. Сравнение результатов работы модели со сверточными и рекуррентными нейросетями на задачах классификации, реферирования текста и машинного перевода. Самостоятельная работа: исследование сфер применения архитектуры «Трансформер» за пределами обработки текстов (анализ временных рядов, обработка изображений, биоинформатика).	2 (лекции) 4 (лабораторные) 2 (СРС)
6	Тема 6. Трансферное обучение Лекция 6. Проблемы недостаточности данных в машинном обучении на основе признаков и в обработке естественного языка. Машинное обучение с	2 (лекции) 4 (лабораторные) 2 (СРС)

	возможностью «переноса» знаний. История развития трансферного обучения. Разновидности трансферного обучения. Преимущества использования архитектуры «Трансформер» для трансфера знаний. Обучение моделей с нуля и тонкая настройка нейросетей. Кодеры, декодеры и автоэнкодеры. Лабораторная работа 1. Знакомство с платформой HuggingFace. Работа с токенизаторами. Выбор модели на хабе. Изучение карточки модели. Тонкая настройка и оценка модели с помощью инструментов HuggingFace. Лабораторная работа 2. Выгрузка и публикация набора данных на HuggingFace Datasets. Самостоятельная работа: использование и публикация простого приложения на HuggingFace Spaces.	
7	Тема 7. Большие языковые модели Лекция 7. Генеративные модели, промпты, техники zero-, one- и few-shot learning. Обучение с подкреплением на основе отзывов людей. Вызовы при использовании и интеграции больших языковых моделей. Гиперпараметры больших языковых моделей. Архитектура Mixture of Experts. Механизм Reasoning. Методы сжатия больших языковых моделей (квантование, дистилляция и проч.). Методы дообучения больших языковых моделей (LoRA, QLoRA, адаптеры). Лабораторная работа 1. Работа с API для получения доступа к большим языковым моделям. Лабораторная работа 2. Разработка простого Telegram-бота на основе больших языковых моделей. Самостоятельная работа: самостоятельное исследование и тестирование методов промпт-	2 (лекции) 4 (лабораторные) 2 (СРС)

	инжиниринга.	
8	Тема 8. Информационная и кибербезопасность в NLP. Перспективы и вызовы использования больших языковых моделей Лекция 8. Retrieval-Augmented Generation. Механизм извлечения информации и генерации текста на основе внешнего источника данных. Методы сбора и хранения данных для систем данного типа. Мультиагентная система. Разработка сервиса на основе систем искусственного интеллекта. Гардрейлы. Методы обеспечения информационной и кибербезопасности систем на основе больших языковых моделей. Лабораторная работа 1. Разработка пайплайна системы Retrieval-Augmented Generation. Лабораторная работа 2. Разработка прототипа сервиса на основе больших языковых моделей. Самостоятельная работа: изучение библиотек LangChain, LlamaIndex, Guardrails AI, Smolagents и других.	2 (лекции) 4 (лабораторные) 2 (СРС)
9	Промежуточная аттестация в формате тестирования и решения практического задания	2

7.3. Учебно-тематический план рабочей программы

№ п/ п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов			
		аудиторных		самостоятельной работы	
		всего, часов	практич еские занятия	всего, часов	практическ ие занятия
1.	Тема 1. Введение в обработку естественного языка	6	4	2	1
2.	Тема 2. Векторные представления текста	6	4	2	1
3.	Тема 3. Сверточные нейронные сети в обработке естественного языка	6	4	2	1
4.	Тема 4. Рекуррентные нейронные сети в обработке естественного языка	6	4	2	1
5.	Тема 5. Механизм внимания и архитектура «Трансформер»	6	4	2	1
6.	Тема 6. Трансферное обучение	6	4	2	1
7.	Тема 7. Большие языковые модели	6	4	2	1
8.	Тема 8. Перспективы и вызовы использования больших языковых моделей	6	4	2	1
9.	Промежуточная аттестация	2			
	Итого	66			

7.4 Контроль и оценка результатов освоения рабочей программы

Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме тестирования и решения практического задания по дисциплине. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

Для текущего и промежуточного контроля создаются фонды оценочных средств (ФОС). ФОС включают в себя педагогические контрольно-измерительные материалы, предназначенные для определения соответствия (или несоответствия) индивидуальных образовательных достижений запланированным образовательным результатам.

7.4.1 Примеры оценочных средств

Примеры тестовых заданий:

Вопрос	Варианты ответов
Какой из методов НЕ используется для построения эмбедингов?	1) Word2Vec 2) TF-IDF 3) Random Forest 4) GloVe
Какой механизм лежит в основе архитектуры «Трансформер»?	1) Сверточные слои 2) Self-Attention 3) Рекуррентные связи 4) Полносвязные сети
Какой формат используется для квантованных языковых моделей в llama.cpp?	1) ONNX 2) GGUF 3) HDF5 4) JSON

Примеры типовых практических заданий:

Сравните методы токенизации:

- 1) Примените токенизацию с помощью NLTK, SpaCy и BERT-токенизатора к одному и тому же тексту.
- 2) Сравните результаты, выделите преимущества и недостатки каждого метода.

Данное практическое задание предполагает выполнение следующих этапов:

- 1) Установка библиотек и настройка окружения.
- 2) Выполнение токенизации с помощью NLTK.
- 3) Выполнение токенизации с помощью SpaCy.
- 4) Составление сравнительной таблицы, выделение таких критериев, как точность токенизации, скорость выполнения функции, анализ контекста, простота использования, сферы применения.

7.5 Учебно-методические материалы для обучающихся и преподавателей

Для подготовки используются учебная литература и комплект материалов, в виде электронных презентаций, содержащие основные определения, терминологический аппарат и иллюстративный материал.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

7.6 Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE). Предусмотрен формат BYOD.

7.7 Информационное обеспечение реализации рабочей программы**Перечень литературы**

1. Jurafsky D., Martin J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. URL: <https://web.stanford.edu/~jurafsky/slp3/> (дата обращения 28.07.2025)
2. Schütze H., Manning C. D., Raghavan P. Introduction to information retrieval. – Cambridge: Cambridge University Press, 2008. – Т. 39. – С. 234-265.
3. Goodfellow I. et al. Deep learning. – Cambridge: MIT press, 2016. – Т. 1. – №. 2.
4. Gao J., Galley M., Li L. Neural approaches to conversational AI //The 41st international ACM SIGIR conference on research & development in information retrieval. – 2018. – С. 1371-1374.

Информационные ресурсы

Образовательный курс на платформе “DeepLearning.AI”. Natural Language Processing Specialization. URL: <https://www.deeplearning.ai/courses/natural-language-processing-specialization/> (дата обращения: 28.07.2025).

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет
«Московский институт электронной техники»

8. Рабочая программа практики

Компьютерная лингвистика

(наименование программы)

Москва 2025

8.1 Область применения рабочей программы

Рабочая программа практики является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Компьютерная лингвистика» и нацелена на получение практических навыков разработки цифровых продуктов, а также на закрепление, углубление и систематизацию знаний, полученных в рамках модулей программы.

Практика является обязательной для прохождения.

Цель практики – получение практических навыков разработки цифровых продуктов, а также на закрепление, углубление и систематизацию знаний и умений, в рамках освоения программы профессиональной переподготовки.

Задачи:

- применение знаний, полученных при изучении дисциплин программы, к решению комплексных практических задач;
- развитие навыков проектной и исследовательской работы;
- отработка умений командного взаимодействия и распределения ролей;
- подготовка и защита итогового проекта.

8.2. Объем практики

Объём практики — 32 ак. часа. Практика организуется с 17 по 37 неделю (распределенно).

8.3 Содержание практики

Практическая подготовка слушателей проходит в индустриальных партнерах ИТ сферы ООО «Нетология».

Компетенции, формируемые в ходе практики:

ID 15 Применяет программные алгоритмы обработки данных для инженерно-научных расчётов.

Индикатор достижения:

Опыт деятельности: анализирует входные данные и выбирает алгоритм их обработки, реализует алгоритм в виде работающего кода, проводит тестирование и оптимизацию решения.

ID 28 Применяет языки программирования для решения профессиональных

задач.

Индикаторы достижения:

Опыт деятельности: использует синтаксис и библиотеки Python/SQL/других языков в соответствии с задачей, интегрирует отдельные программные модули в единое приложение, оформляет код с комментариями.

ID 32 Использует СУБД при разработке ПО.

Индикаторы достижения:

Опыт деятельности: разработки и оптимизации запросов; работы с реляционными и NoSQL СУБД; взаимодействия СУБД с внешними приложениями; проектирования структур данных.

ID 37 Применяет искусственный интеллект и машинное обучение.

Индикаторы достижения:

Опыт деятельности: выбирает готовую модель ИНС или разрабатывает собственную, обрабатывает данные для обучения/тестирования модели, оценивает качество работы модели по метрикам.

На практике слушателям предлагается групповое задание по теме, выданной на месте практики и согласованной с руководителем практики. Разработка осуществляется командой слушателей и курируется представителем с места практики. При прохождении практики слушатели развивают приобретенные навыки, знакомятся с имеющимися технологическими решениями, изучают инструкции по работе действующих специалистов, самостоятельно проводят поиск научно-технической информации по тематике проекта, планируют этапы выполнения проекта, учатся оформлять и утверждать техническое задание. В ходе прохождения практической подготовки слушателям необходимо разработать цифровой продукт по согласованной теме.

Этапы прохождения практики

Этап	Содержание	Часы, ак.ч.	Самостоятельная работа
1	Введение в проект, формирование команд,	8	Анализ ТЗ, поиск аналогов, изучение инструментов

	распределение ролей, постановка задач		
2	Разработка прототипа решения (обработка данных, проектирование алгоритмов)	10	Подготовка и тестирование отдельных модулей
3	Интеграция и отладка решения	8	Подготовка итоговой презентации и демонстрации
4	Подготовка отчетной документации	6	Итоговая проверка документации, репетиция презентации
Итого	32		

8.4. Пример типового задания по практике

Задача: разработать алгоритм автоматического выявления и классификации метафор в художественном тексте и реализовать скрипт автоматической проверки орфографии и грамматики в текстовых документах.

Описание выполнения:

1. Выбрать корпус текстов
2. Разработать критерии определения метафор.
3. Составить и разметить небольшой обучающий набор.
4. Протестировать несколько методов (ключевые слова, синтаксический анализ, простые ML-модели).
5. Подготовить аналитический отчёт о типах метафор и частотности их использования.
6. Определить язык(и) проверки (русский, английский).
7. Использовать готовые библиотеки (py morphology2, LanguageTool и др.).
8. Реализовать обработку загруженного файла и выдачу отчёта с найденными ошибками.
9. Настроить возможность исправления ошибок в интерфейсе.

Этапы выполнения:

№	Этап практики	Задачи и содержание	Отчетные документы
1	Сбор и предобработка данных	Собрать корпус текстов из открытых источников; выполнить очистку и нормализацию текстов (токенизация, удаление стоп-слов, лемматизация).	Подготовить отчёт о выбранном источнике данных и способах очистки; изучить лучшие практики по подготовке текстовых корпусов.
2	Морфологический и синтаксический анализ	Использовать инструменты (Mystem, SpaCy) для морфологической разметки; построить деревья зависимостей для выбранных предложений; интерпретировать полученные структуры.	Проанализировать ошибки морфологического и синтаксического анализаторов; подготовить отчет о наборе правил для исправления наиболее частотных ошибок.
3	Моделирование и оценка	Построить модели классификации и регрессии для решения выбранной задачи; провести кросс-валидацию; выбрать метрики оценки; визуализировать результаты.	Изучить примеры успешных моделей для подобной задачи; попробовать различные алгоритмы (SVM, логистическая регрессия, деревья решений) и сравнить их эффективность. Подготовить отчет о сравнении.
4	Итоговый проект	в командном режиме собрать все этапы в единую цепочку, доработать модель, подготовить презентацию; представить результаты экспертному жюри.	Оформить презентацию; подготовить репозиторий проекта (н-р, GitHub), обеспечив воспроизводимость.

Результат: рабочий прототип с консольным или простым веб-интерфейсом, инструкция по установке и использованию, отчеты выполнения этапов.

8.4.1 Критерии оценки результата

Основной формой итоговой оценки является защита проекта (итоговая аттестация). Каждая команда представляет готовый продукт, демонстрирует ход решения задачи, результаты моделирования, а также отвечает на вопросы комиссии. В процессе выполнения практики ведётся текущий контроль: проверяются отчёты по каждому этапу, проводится ревью кода и документации.

8.5. Учебно-методические материалы для обучающихся

Для выполнения индивидуального задания в рамках практики используется учебная литература и примеры выполнения практических заданий, представленные в соответствующих разделах учебных дисциплин, в рамках которых происходит обучение подзадачам, а также документация и руководства по библиотекам NLP (SpaCy, NLTK, Stanza), открытые наборы данных для задач NLP (Kaggle, OpenSubtitles).

8.6. Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, LibreOffice, браузер (Яндекс и пр.), среда разработки (Jupyter, Notepad++, VS CODE), Git, Gitlab, Gitverse, виртуальная машина/контейнер (Virtual Box, Docker) с установленной debian-based операционной системой (ОС AstraLinux, ОС Ubuntu), интерпретатор языка Python, системы визуализации данных PlantUML, база данных PostgreSQL. Используется формат BYOD.

8.7 Информационное обеспечение реализации рабочей программы

Список литературы

- 1) Джурафски Д., Мартин Дж. Х. Обработка речи и языка : учебник / пер. с англ. — 3-е изд. — Прентис-Холл, 2023. — 1024 с.
- 2) Маннинг К., Шютце Х. Основы статистической обработки естественного языка : учебное пособие / пер. с англ. — Обновл. изд. — MIT Press, 2020. — 680 с.
- 3) Бёрд С., Кляйн Э., Лопер Э. Обработка естественного языка с помощью Python : практическое руководство / пер. с англ. — 2-е изд. — O'Reilly Media, 2021. — 504 с.
- 4) Шапиро Б., Ульрих К. Проектирование пользовательских интерфейсов для науки о данных / пер. с англ. — Apress, 2022. — 320 с.

5) Камбрия Э., Пория С., Хазарика Д., Квок К. Анализ тональности: от теории к практике / пер. с англ. — Springer, 2022. — 260 с.

6) Аггарвал Ч. Машинное обучение для работы с текстом / пер. с англ. — Springer, 2022. — 350 с.

7) Катценбах Дж. Р., Смит Д. К. Мудрость команд: создание высокоэффективной организации / пер. с англ. — Обновл. изд. — HarperBusiness, 2021. — 320 с.

VI. Аттестация по Программе

После завершения обучения по Программе обучающиеся допускаются к итоговой аттестации.

Аттестация проводится с участием представителей профильных индустриальных партнеров в форме демонстрационного экзамена и предусматривает выполнение обучающимся профессиональных задач и оценку результатов и (или) процесса выполнения – проверку сформированности в рамках Программы цифровых компетенций.

Задания демонстрационного экзамена разрабатываются с участием организаций-работодателей, отраслевых партнеров и профессиональных сообществ.

VI.1.1. Форма итоговой аттестации

Демонстрационный экзамен в формате комплексного практического задания (кейса) или хакатона и презентационной защиты.

VI.1.2. Общие принципы

Демонстрационный экзамен моделирует выполнение проекта в условиях, приближенных к профессиональной деятельности. Участники работают в командах (3–5 человек). Результатом является готовый программный продукт/прототип, сопровождаемый документацией и устной защитой.

VI.1.3. Примеры формулировки заданий

1) На основании предоставленного набора данных (тексты новостных статей и метаданные) спроектировать и реализовать систему автоматической классификации текстов по тематикам.

В рамках выполнения задания необходимо:

1. Предварительно обработать данные (токенизация, лемматизация, удаление стоп-слов).
2. Сохранить предобработанные данные в реляционную БД (PostgreSQL) с корректной структурой и индексами.
3. Разработать Python-скрипт, который выполняет анализ текста и классифицирует его по заданным категориям.
4. Для хранения метаданных эксперимента использовать NoSQL БД (MongoDB).
5. Подготовить визуализацию результатов и итоговую презентацию.

2) На основании предоставленного корпуса отзывов о товарах/услугах на русском языке разработать систему автоматического анализа тональности (позитив, негатив, нейтрально) с использованием инструментов обработки текста, баз данных и визуализации.

В рамках выполнения задания требуется:

1. Предобработка текста

Очистка текста от лишних символов и HTML-тегов. Токенизация, лемматизация, удаление стоп-слов.

2. Хранение данных

Проектирование структуры реляционной БД для хранения исходных и обработанных данных. Загрузка данных в PostgreSQL.

3. Разработка анализа тональности

Использование Python-библиотек (например, NLTK, spaCy, DeepPavlov) для классификации отзывов. Определение тональности с учётом ключевых слов и/или модели машинного обучения.

4. NoSQL-хранилище

Сохранение статистики по тональности и метаданных эксперимента в MongoDB.

5. Визуализация и защита

Построение диаграмм (доля позитивных/негативных отзывов). Презентация алгоритма, результатов и выводов комиссии.

3) Разработать прототип поисковой системы, позволяющей осуществлять тематический поиск по базе научных статей с ранжированием результатов по релевантности.

В рамках выполнения задания требуется:

1. Загрузка и подготовка данных. Получение корпуса статей в формате JSON/CSV. Нормализация текста, извлечение ключевых слов и фраз.
2. Проектирование базы данных. Создание структуры реляционной БД (PostgreSQL) для хранения метаданных (название, автор, ключевые слова, аннотация). Хранение полнотекстового индекса.
3. Разработка поискового алгоритма. Реализация полнотекстового поиска средствами PostgreSQL (tsvector, tsquery) или Elasticsearch (по желанию). Сортировка результатов по релевантности.
4. NoSQL-хранилище. Сохранение истории поисковых запросов пользователей в MongoDB.
5. Интерфейс и защита. Создание консольного или веб-интерфейса для ввода запроса и вывода результатов. Подготовка презентации к защите.

4) Автоматическая генерация резюме текста научных статей

В рамках выполнения задания требуется:

1. Загрузка и предобработка данных: получить корпус статей (в формате PDF/JSON/CSV), выделить заголовок, аннотацию и основной текст. Очистить текст от лишних символов, провести токенизацию, лемматизацию, удаление стоп-слов.
2. Хранение данных: спроектировать структуру реляционной БД (PostgreSQL) для хранения исходных и обработанных данных, включая текст статей и сгенерированные краткие резюме.

3. Разработка алгоритма: применить Python-библиотеки для генерации кратких резюме (например, gensim, HuggingFace Transformers). Реализовать как экстрактивное, так и абстрактное суммирование.
4. NoSQL-хранилище: сохранять метрики качества (ROUGE, BLEU) и историю экспериментов в MongoDB.
5. Визуализация и защита: подготовить таблицы и графики с сравнением качества разных алгоритмов, представить результаты в презентации.

5) Автоматическое выявление фейковых новостей

В рамках выполнения задания требуется:

1. Подготовка корпуса данных: загрузить набор новостных статей (CSV/JSON), помеченных как достоверные и фейковые. Провести очистку текста (удаление HTML, спецсимволов), токенизацию и лемматизацию.
2. Хранение данных: разработать схему PostgreSQL для хранения исходных текстов, признаков и результатов классификации.
3. Модель машинного обучения: реализовать классификатор (Logistic Regression, SVM, BERT) для автоматического определения фейковых новостей.
4. NoSQL-хранилище: сохранять результаты классификации и логи экспериментов в MongoDB.
5. Визуализация и защита: построить графики точности (Precision, Recall, F1-score), разработать дашборд с результатами, подготовить презентацию для защиты.

VI.1.3. Проверяемые компетенции и критерии оценки

№	Критерий	Показатели результата	Компетенции	Вес критерия (макс. балл)	Высокий уровень (8-10б.)	Базовый уровень (5-7б.)	Низкий уровень (1-4б.)
1	Анализ данных и постановка задачи	Определяет цели, формулирует требования, выбирает методы обработки	ID 15	15	Цели и задачи чёткие, методы обоснованы	Цели и задачи определены, методы частично обоснованы	Цели нечеткие, методы случайны
2	Реализация алгоритмов	Пишет рабочий и оптимизированный код, реализует алгоритмы	ID 15, ID 28	15	Код рабочий, оптимизирован, документирован	Код рабочий, но есть избыточность	Код с ошибками, неполный функционал
3	Применение языков программирования	Использует Python и другие инструменты для решения задач	ID 28	15	Языки применяются эффективно, библиотеки используются по назначению	Применение с частичными ошибками	Конструкции неверны, результата нет
4	Работа с базами данных (SQL/NoSQL)	Создаёт структуру БД, формирует корректные запросы	ID 32	15	Структура корректна, запросы оптимальны	Структура верна, но неоптимальна	Структура и запросы некорректны
5	Применение ИИ и машинного обучения	Использует готовые модели или разрабатывает собственные	ID 37	15	Модель выбрана/разработана верно, результаты воспроизводимы	Модель верна, но результаты неполные	Модель неверна или не реализована
6	Качество анализа и интерпретации	Применяет корректные методы, делает выводы	ID 15, ID 37	15	Методы и выводы верны	Методы верны, выводы частично обоснованы	Методы неверны, выводы отсутствуют
8	Презентация проекта	Представляет результаты, аргументирует решения		10	Презентация логична, ответы уверенные	Презентация понятна, ответы частично верны	Презентация бессистемна, ответы отсутствуют

Итого: 100 баллов. Менее 50 баллов соответствует оценке «неудовлетворительно», 51-69 баллов соответствуют оценке

«удовлетворительно», 70 – 85 баллов оценке «хорошо», 86-100 баллов оценке «отлично».

VII. Завершение обучения по Программе

Лицам, завершившим обучение по Программе и достигших целевого уровня сформированности цифровых компетенций по результатам итоговой оценки и прошедших итоговую аттестацию, присваивается дополнительная ИТ-квалификация, установленная Программой.

При освоении Программы параллельно с получением высшего образования диплом о профессиональной переподготовке выдается не ранее получения соответствующего документа об образовании и о квалификации (за исключением лиц, имеющих среднее профессиональное или высшее образование).

Лицам, не прошедшим итоговую аттестацию или получившим на итоговой аттестации неудовлетворительные результаты, а также лицам, освоившим часть Программы и (или) отчисленным из образовательной организации высшего образования, реализующей Программу, выдается справка об обучении или о периоде обучения.

Согласовано:

Директор Института ЛПО



М.Г. Евдокимова

Руководитель проекта
«Цифровые кафедры»



В.В. Кокин