



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«МИРЭА – Российский технологический университет»

Институт кибербезопасности и цифровых технологий

УТВЕРЖДАЮ

Директор ИКБ

Бакаев А.А.

«__» _____ 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Москва
2022

Рабочая программа дисциплины (модуля) составлена на основании Федерального государственного образовательного стандарта высшего образования

- бакалавриат по направлению подготовки 01.03.02 Прикладная математика и информатика, утвержденного приказом Министерства образования и науки Российской Федерации №9 от 10.01.2018. Зарегистрировано в Минюсте России 06.02.2018 № 49937

Составители:

кандидат технических наук, доцент

В. Г. Литвинов

Заведующий
кафедрой технической кибернетики

доктор
технических наук,
доцент

Рабочая программа обсуждена на заседании кафедры технической кибернетики.
Протокол №5 от 23.11.2021.

Руководитель основной профессиональной образовательной программы высшего образования:
Искусственный интеллект и компьютерные науки по направлению подготовки 01.03.02

Э. И. Коломиец

1. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

1.1. Цели и задачи изучения дисциплины (модуля)

Цель дисциплины: ознакомление студентов с основными моделями, методами и подходами компьютерной лингвистики.

Задачи дисциплины:

1. Сформировать у студентов теоретические знания и практические навыки в области решения задач компьютерной лингвистики с использованием современных инструментов.
2. Ознакомление студентов с основными методами, применяемыми при решении задач компьютерной лингвистики.
3. Выработка у студентов приёмов и навыков решения конкретных задач.
4. Ознакомление студентов с современными подходами к решению задач компьютерной лингвистики.

1.2 Перечень формируемых компетенций и индикаторы их достижения, требования к уровню подготовки обучающегося, завершившего изучение данной дисциплины (модуля)

Планируемые результаты освоения образовательной программы (компетенции обучающихся) определяются требованиями стандарта по направлению подготовки (специальности) и формируются в соответствии с матрицей компетенций образовательной программы.

Планируемые результаты обучения по дисциплине (модулю) - знания, умения, навыки и (или) опыт деятельности формируются в соответствии с индикаторами достижения компетенций и результатами освоения образовательной программы (таблица 1).

Таблица 1

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине (модулю)
ПК-2 Способен использовать инструментальные средства для решения задач машинного обучения	ПК-2.3 Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения	Знает классификацию существующих моделей, используемых при обработке естественных языков. Умеет применять и модифицировать модели при решении задач компьютерной лингвистики. Владеет навыками использования различных методов и инструментов для решения задач компьютерной лингвистики.;
ПК-4 Способен классифицировать и идентифицировать задачи искусственного интеллекта,	ПК-4.1 Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от	Знает классификацию существующих моделей, используемых при обработке естественных языков. Умеет применять и модифицировать модели при решении задач компьютерной лингвистики. Владеет навыками использования различных методов и инструментов для решения задач

выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	особенностей проблемной и предметной областей; ПК-4.2 Выбирает методы и инструментальные средства искусственного интеллекта для решения задач в зависимости от особенностей проблемной и предметной областей;	компьютерной лингвистики.; Знает основные модели представления слов. Умеет выбирать подходящие инструменты для решения поставленных задач. Владеет навыками подбора оптимальных методов для решения задач компьютерной лингвистики.;
ПК-5 Способен разрабатывать и применять методы машинного обучения для решения задач	ПК-5.1 Проводит анализ требований и определяет необходимые классы задач машинного обучения; ПК-5.2 Определяет метрики оценки результатов моделирования и критерии качества построенных моделей; ПК-5.3 Принимает участие в оценке, выборе и при необходимости разработке методов машинного обучения;	Знает основные методы машинного обучения, применяемые при решении задач компьютерной лингвистики. Умеет выявлять и формулировать требования для поставленной задачи. Владеет навыками применения различных методов машинного обучения в задачах компьютерной лингвистики.; Знает метрики, используемые при решении задач компьютерной лингвистики. Умеет определять качество решения задач построенной моделью. Владеет навыками оценки качества моделей.; Знает преимущества и недостатки моделей, используемых при решении задач компьютерной лингвистики. Умеет выбирать наиболее адекватную поставленной задаче модель. Владеет навыками применения различных моделей для решения одной задачи.;
ПК-6 Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	ПК-6.1 Осуществляет оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи; ПК-6.2 Разрабатывает системы искусственного интеллекта на основе моделей	Знает основные методы машинного обучения, применяемые при решении задач компьютерной лингвистики. Умеет выявлять и формулировать требования для поставленной задачи. Владеет навыками применения различных методов машинного обучения в задачах компьютерной лингвистики.; Знает метрики, используемые при решении задач компьютерной лингвистики. Умеет определять качество решения задач построенной моделью. Владеет навыками оценки качества моделей.

	искусственных нейронных сетей и инструментальных средств	
--	---	--

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Перечень предшествующих и последующих дисциплин, формирующих общекультурные и профессиональные компетенции (таблица 2)

Таблица 2

№	Код и наименование компетенции	Предшествующие дисциплины (модули)	Последующие дисциплины (модули)
1	ПК-2 Способен использовать инструментальные средства для решения задач машинного обучения	Технологическая (научно-технологическая) практика, Машинное обучение, Искусственный интеллект, Машинное зрение, Введение в социальное и гуманитарное знание, Математическое моделирование в гуманитарных науках, Технологии программирования на Python, Философия искусственного интеллекта, Научно-исследовательская работа, Инструментальные средства решения научных задач, Введение в специальность	Искусственный интеллект, Машинное зрение, Философия искусственного интеллекта, Научно-исследовательская работа, Выполнение и защита выпускной квалификационной работы, Инструментальные средства решения научных задач
2	ПК-4 Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	Технологическая (научно-технологическая) практика, Машинное обучение, Искусственный интеллект, Машинное зрение, Математическое моделирование в гуманитарных науках, Технологии программирования на Python, Философия искусственного интеллекта, Научно-исследовательская работа, Инструментальные средства решения научных задач, Введение в специальность	Искусственный интеллект, Машинное зрение, Философия искусственного интеллекта, Научно-исследовательская работа, Выполнение и защита выпускной квалификационной работы, Инструментальные средства решения научных задач
3	ПК-5 Способен разрабатывать и применять методы машинного обучения для решения задач	Технологическая (научно-технологическая) практика, Машинное обучение, Искусственный интеллект, Машинное зрение, Введение в социальное и гуманитарное знание, Математическое	Искусственный интеллект, Машинное зрение, Научно-исследовательская работа, Выполнение и защита выпускной квалификационной работы

		моделирование в гуманитарных науках, Технологии программирования на Python, Научно-исследовательская работа	
4	ПК-6 Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	Анализ социальных сетей, Машинное обучение, Машинное зрение, Математическое моделирование в гуманитарных науках, Научно-исследовательская работа, Анализ больших данных	Анализ социальных сетей, Машинное зрение, Научно-исследовательская работа, Выполнение и защита выпускной квалификационной работы, Анализ больших данных

3. ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ) С УКАЗАНИЕМ ОБЪЕМА КОНТАКТНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ С ПРЕПОДАВАТЕЛЕМ (ПО ВИДАМ УЧЕБНЫХ ЗАНЯТИЙ) И ОБЪЕМА САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ, А ТАКЖЕ СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ), СТРУКТУРИРОВАННОЕ ПО ТЕМАМ (РАЗДЕЛАМ) С УКАЗАНИЕМ ОБЪЕМА ОТВЕДЕННОГО НА НИХ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ

Таблица 3

<u>Объём дисциплины: 4 ЗЕТ</u>
<u>Восьмой семестр</u>
Объём контактной работы: 50 час.
Лекционная нагрузка: 24 час.
<i>Активные и интерактивные</i>
Лекция 1. Формальный язык. Виды грамматик (иерархия Хомского). Регулярные и контекстно-зависимые языки, их отличия. (2 час.)
Лекция 2. Языковые модели. Модель непрерывного мешка слов (CBoW). (2 час.)
Лекция 3. Архитектура Skip-Gram. (2 час.)
Лекция 4. Глобальные векторы представления слов (GloVe). (2 час.)
Лекция 5. Анализ зависимостей. (2 час.)
Лекция 6. Рекуррентные нейронные сети. RNN модель машинного перевода. (2 час.)
Лекция 7. Управляемые рекуррентные блоки. Долгая краткосрочная память. (2 час.)
Лекция 8. Машинный перевод Seq2Seq. (2 час.)
Лекция 9. Механизм внимания. (2 час.)
Лекция 10. Декодер модели последовательности. (2 час.)
Лекция 11. Оценка качества систем машинного перевода. Генерация, оценка и метрики (GEM)
Лекция 12. Обработка неизвестных слов. Словесные и символьные модели. (2 час.)
Лабораторные работы: 24 час.
<i>Активные и интерактивные</i>
Лабораторная работа 1. Предварительная обработка и частотный анализ текста. (6 час.)
Лабораторная работа 2. Анализ тональности. (6 час.)
Лабораторная работа 3. Векторные представления слов. (6 час.)
Лабораторная работа 4. Машинный перевод. (6 час.)
Контролируемая аудиторная самостоятельная работа: 2 час.
<i>Активные и интерактивные</i>
Подготовка к экзамену. (2 час.)
Самостоятельная работа: 58 час.
<i>Активные и интерактивные</i>
Изучение теоретического материала и подготовка к лабораторным работам. (58 час.)
Контроль (Экзамен) (36 час.)

4. ПЕРЕЧЕНЬ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ И ИННОВАЦИОННЫХ МЕТОДОВ ОБУЧЕНИЯ, ИСПОЛЬЗУЕМЫХ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО

Инновационные методики

1. Использование мультимедийного оборудования при проведении лекционных занятий.
2. Использование на лекциях интерактивных презентаций, созданных в системе Jupyter Notebook.

5. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ (В ТОМ ЧИСЛЕ ОТЕЧЕСТВЕННОГО ПРОИЗВОДСТВА), НЕОБХОДИМОЕ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

5.1 Требования к материально-техническому обеспечению

Табл

№ п/п	Тип помещения	Состав оборудования и технических средств обучения
1	Аудитория для проведения лекционных занятий	Учебная аудитория для проведения занятий лекционного типа, оборудованная учебной мебелью: столы, стулья для обучающихся; стол, стул для преподавателя; набором демонстрационного оборудования и учебно-наглядных пособий; ноутбуком с выходом в сеть Интернет, проектором; экраном настенным;
2	Аудитория для проведения лабораторных работ	учебная аудитория для проведения лабораторных работ, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук с выходом в сеть Интернет), компьютерами со специализированным программным обеспечением; учебной мебелью: столы, стулья для обучающихся; стол, стул для преподавателя.
3	Аудитория для самостоятельной работы	Помещение для самостоятельной работы, оснащенное компьютерами с доступом в Интернет и в электронно-информационную образовательную среду Самарского университета.
4	Аудитория для выполнения контролируемой аудиторной	Учебная аудитория, оборудованная учебной мебелью: столы, стулья для обучающихся; стол,
5	Аудитория для проведения, текущего контроля и промежуточной аттестации	Учебная аудитория для проведения текущего контроля и промежуточной аттестации, оборудованная учебной мебелью: столы, стулья для обучающихся; стол, стул для преподавателя;

5.2 Перечень лицензионного программного обеспечения

1. MS Windows 10 (Microsoft)

в том числе перечень лицензионного программного обеспечения отечественного производства:

1. Kaspersky Endpoint Security (Kaspersky Lab)

5.3 Перечень свободно распространяемого программного обеспечения

1. Python

в том числе перечень свободно распространяемого программного обеспечения отечественного

1. Яндекс.Браузер

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Основная литература

1. Короткова, М. А. Задачник по курсу "Математическая лингвистика и теория автоматов : учебное пособие / М. А. Короткова, Е. Е. Трифонова. — Москва : НИЯУ МИФИ, 2012. — 92 с. — ISBN 978-5-7262-1702-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/75843> (дата обращения: 00.00.0000). — Режим доступа: для
2. Чигарина, Е. И. Теория формальных языков и грамматик [Электронный ресурс] : [учеб. пособие]. - Самара.: Изд-во Самар. ун-та, 2018. - on-line
3. Филиппов, Ф. В. Моделирование нейронных сетей глубокого обучения : учебное пособие / Ф. В. Филиппов. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2019. — 79 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/180053> (дата обращения: 00.00.0000). — Режим доступа: для авториз.

6.2. Дополнительная литература. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

1. Полупанов, Д. В. Программирование в Python 3 : учебное пособие / Д. В. Полупанов, С. Р. Абдюшева, А. М. Ефимов. — Уфа : БашГУ, 2020. — 164 с. — ISBN 978-5-7477-5230-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/179915> (дата обращения: 00.00.0000). — Режим доступа: для авториз.

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

Табл

№ п/п	Наименование ресурса	Адрес	Тип доступа
1	Python	python.org	Открытый
2	Natural Language Toolkit	nltk.org	Открытый
3	Научная электронная библиотека «КиберЛенинка»	https://cyberleninka.ru	Открытый ресурс
4	Архив научных журналов на платформе НЭИКОН	https://archive.neicon.ru/xmlui/	Открытый ресурс

6.4 Перечень информационных справочных систем и профессиональных баз данных, необходимых для освоения дисциплины (модуля)

6.4.1 Перечень информационных справочных систем, необходимых для освоения дисциплины

Табл

№ п/п	Наименование информационного ресурса	Тип и реквизиты ресурса
1	СПС КонсультантПлюс	Информационная справочная система, 2020_12_29_д_ЭК-112-20
2	Система интегрированного поиска EBSCO Discovery Service EBSCO	Информационная справочная система, Сублицензионный договор №156-EBSCO-21 от

6.4.2 Перечень современных профессиональных баз данных, необходимых для освоения

Табл

№ п/п	Наименование информационного ресурса	Тип и реквизиты ресурса
1	Национальная электронная библиотека ФГБУ "РГБ"	Профессиональная база данных, Договор № 101/НЭБ/4604 от 13.07.2018
2	Электронно-библиотечная система elibrary (журналы)	Профессиональная база данных, Договор № SU-01-10/2021 на оказание услуг доступа к электронным изданиям от 22.10.2021, Лицензионное соглашение № 953 от 26.01.2004
3	Scopus издательской корпорации Elsevier	Профессиональная база данных, Заявление-21-1702-01024

6.5 ИСПОЛЬЗОВАНИЕ ЭЛЕКТРОННОЙ ИНФОРМАЦИОННО-ОБРАЗОВАТЕЛЬНОЙ СРЕДЫ, ЭЛЕКТРОННЫХ БИБЛИОТЕЧНЫХ СИСТЕМ, ЭЛЕКТРОННОГО ОБУЧЕНИЯ И ДИСТАНЦИОННЫХ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ ДЛЯ ОСВОЕНИЯ

В процессе освоения дисциплины (модуля) обучающиеся обеспечены доступом к электронной информационно-образовательной среде и электронно-библиотечным системам (<http://lib.ssau.ru/els>). В процессе освоения дисциплины (модуля) могут применяться

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Лекция представляет собой систематическое устное изложение учебного материала. С учетом целей и места в учебном процессе различают лекции вводные, установочные, текущие, обзорные и заключительные. В зависимости от способа проведения выделяют лекции:

- информационные;
- проблемные;
- визуальные;
- лекции-конференции;
- лекции-консультации;
- лекции-беседы;
- лекция с эвристическими элементами;
- лекция с элементами обратной связи.

По данной дисциплине применяются в основном информационные лекции - проводятся с использованием объяснительно иллюстративного метода изложения; это традиционный для высшей школы тип лекций.

Также применяются лекции с элементами обратной связи. В данном случае подразумевается изложение учебного материала и использование знаний по изученному ранее учебному материалу. Обратная связь устанавливается посредством ответов обучающихся на вопросы преподавателя по ходу лекции. Чтобы определить осведомленность обучающихся по излагаемой проблеме, в начале какого-либо раздела лекции задаются необходимые вопросы. Если обучающиеся правильно отвечают на вводный вопрос, преподаватель может ограничиться кратким тезисом или выводом и перейти к следующему вопросу.

Лабораторные работы - форма организации обучения, которая направлена на формирование практических умений и навыков и является связующим звеном между самостоятельным теоретическим освоением студентами учебной дисциплины и применением ее положений на практике. Лабораторные работы проводятся в целях: выработки практических умений и приобретения навыков в обработке реальных данных с помощью автоматизированной информационной системы специального назначения, интерпретация полученных результатов, построении случайных функциональных характеристик с помощью автоматизированной информационной системы, построение и реализация математических моделей. Главным их содержанием является практическая работа каждого студента. Подготовка студентов к лабораторной работе и её выполнение осуществляется по заданию, которое разрабатывается преподавателем и доводится до обучающихся до начала проведения лабораторной работы. Лабораторные работы составляют значительную часть всего объема аудиторных занятий и имеют важнейшее значение для усвоения программного материала.

Самостоятельная работа студентов является одной из важнейших составляющих учебного процесса, в ходе которого происходит формирование знаний, умений и навыков в учебной, научно-исследовательской, профессиональной деятельности, формирование профессиональных компетенций будущего магистра.

Рабочей программой дисциплины предусмотрены следующие виды самостоятельной работы студентов:

- самостоятельная работа, обеспечивающая подготовку к текущим лабораторным занятиям;
- самостоятельная работа, обеспечивающая подготовку к экзамену.

Следует выделить подготовку к экзамену как особый вид самостоятельной работы. Основное его отличие от других видов самостоятельной работы состоит в том, что обучающиеся решают задачу систематизации учебного материала, применения приобретенных знаний и умений в качестве структурных элементов компетенций, формирование которых выступает целью и результатом освоения образовательной программы.