Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 22.05.2024 17:54:30
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

**(МОСКОВСКИЙ ПОЛИТЕХ)**

Факультет информационных технологий

УТВЕРЖДАЮ
Декан факультета
«Информационные технологии»
_____ / Д.Г.Демидов /
«15» февраля 2024г.

# РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

## «Обработка естественного языка и машинный анализ текстов»

Направление подготовки
**09.03.01 Информатика и вычислительная техника**

Профиль
**«Системная и программная инженерия»**

Квалификация
**Бакалавр**

Формы обучения
**очная**

Москва, 2024 г.

**Разработчик(и):**

старший преподаватель кафедры «Информационная безопасность»

/А.Ю. Гневшев/

**Согласовано:**
И.о. заведующего кафедрой «Информационная безопасность»,

/А.Ю. Гневшев/

Руководитель образовательной программы,

/А.Ю. Гневшев/

Заведующий кафедрой «Инфокогнитивные технологии»,

доцент, к.т.н. /Е.А.Пухова/

# Содержание

1 Цели, задачи и планируемые результаты обучения по дисциплине ....4
2 Место дисциплины в структуре образовательной программы ....4
3 Структура и содержание дисциплины ....5
3.1 Виды учебной работы и трудоемкость ....5
3.2 Тематический план изучения дисциплины ....5
3.3 Содержание дисциплины ....6
4 Учебно-методическое и информационное обеспечение ....6
4.1 Основная литература ....6
4.2 Дополнительная литература ....7
4.3 Электронные образовательные ресурсы ....7
5 Материально-техническое обеспечение ....7
5.1 Требования к оборудованию и помещению для занятий ....7
5.2 Требования к программному обеспечению ....7
6 Методические рекомендации ....7
6.1 Методические рекомендации для преподавателя по организации обучения ....7
6.2 Методические указания для обучающихся по освоению дисциплины ....8
7 Фонд оценочных средств ....8
7.1 Методы контроля и оценивания результатов обучения ....8
7.2 Шкала и критерии оценивания результатов обучения ....8
7.3 Оценочные средства ....11
7.3.1 Экзаменационные вопросы ....11
7.3.2 Типовые практические задания ....12
7.3.3 Типовой экзаменационный билет ....13

## 1 Цели, задачи и планируемые результаты обучения по дисциплине

Цель изучения дисциплины – изучение методов и алгоритмов анализа и обработки текстов на естественном языке с помощью компьютерных технологий.

Обучение по дисциплине «Обработка естественного языка и машинный анализ текстов» направлено на формирование у обучающихся следующих компетенций:

| Код и наименование компетенций | Индикаторы достижения компетенции |
|---|---|
| ОПК-8. Способен разрабатывать алгоритмы и программы, пригодные для практического применения | ИОПК-8.1. Знает основные технологии создания и внедрения информационных систем, стандарты управления жизненным циклом информационной системы.<br>ИОПК-8.2. Умеет проектировать блок-схемы алгоритмов, оценивать производительность алгоритмов и затраты памяти на работу алгоритма, разрабатывать программы на основе спроектированного алгоритма и проводить отладку программы, применять методы системного анализа и математического моделирования при разработке и эксплуатации ИС, проводить структурный анализ, функциональный анализ, объектно-ориентированный анализ иерархии классов. осуществлять организационное обеспечение выполнения работ на всех стадиях и в процессах жизненного цикла информационной системы.<br>ИОПК-8.3. Владеет навыками разработки программ, построения блок-схем алгоритмов и оценки производительности алгоритмов, работы с унифицированным языком визуального моделирования, составления плановой и отчетной документации по управлению проектами создания информационных систем на стадиях жизненного цикла. |

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются поэтапно в ходе освоения обучающимися дисциплин (модулей), практик в соответствии с учебным планом и календарным графиком учебного процесса.

## 2 Место дисциплины в структуре образовательной программы

Дисциплина относится к базовой части основной профессиональной образовательной программы бакалавриата.

Дисциплина взаимосвязана логически и содержательно-методически со следующими дисциплинами и практиками ОПОП:

- Разработка технических текстов и документации.

## 3 Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетных единицы, т.е. 144 академических часа (из них 72 часа – самостоятельная работа студентов и 72 часа – аудиторные занятия).

Разделы дисциплины изучаются в седьмом семестре обучения, т.е. на четвертом курсе. Форма контроля – курсовой проект, экзамен.

### 3.1 Виды учебной работы и трудоемкость (по формам обучения)

3.1.1 Очная форма обучения

| № п/п | Вид учебной работы | Количество часов | Семестры | |
|---|---|---|---|---|
| | | | 7 | |
| **1** | **Аудиторные занятия** | **72** | 72 | |
| | В том числе: | | | |
| 1.1 | Лекции | 4 | 4 | |
| 1.2 | Семинарские/практические занятия | | | |
| 1.3 | Лабораторные занятия | 68 | 68 | |
| **2** | **Самостоятельная работа** | **72** | 72 | |
| **3** | **Промежуточная аттестация** | | | |
| | Экзамен, зачет, курсовые работы | | Экзамен, курсовая работа | |
| | Итого: | **144** | 144 | |

### 3.2 Тематический план изучения дисциплины (по формам обучения)

3.2.1 Очная форма обучения

| № п/п | Разделы/темы дисциплины | Трудоемкость, час | | | | | |
|---|---|---|---|---|---|---|---|
| | | Всего | Аудиторная работа | | | | Самостоятельная работа |
| | | | Лекции | Семинарские/ практические занятия | Лабораторные занятия | Практическая подготовка | |
| 1 | Основы лингвистики | 36 | 2 | | 16 | | 18 |
| 2 | Алгоритмы машинного обучения для обработки естественных языков | 36 | 2 | | 16 | | 18 |
| 3 | Классификация и кластеризация текстов | 36 | | | 18 | | 18 |
| 4 | Создание систем обработки естественного языка | 36 | | | 18 | | 18 |
| **Итого** | | **144** | **4** | | **68** | | **72** |

### 3.3 Содержание дисциплины

1. Введение в дисциплину “Обработка естественного языка и машинный анализ текстов”.
2. Основы лингвистики и теории информации.
3. Введение в алгоритмы и методы машинного обучения.
4. Классификация текстов и методы их кластеризации.
5. Извлечение информации из текстовых данных.
6. Распознавание языка и анализ тональности текста.
7. Построение рекомендательных систем на основе анализа текстов.
8. Создание систем автоматического перевода и аннотирования текстов.
9. Методы обработки естественного языка в задачах информационного поиска.
10. Разработка чат-ботов и систем разговорного интерфейса.
11. Обработка текстов на изображениях и оптическое распознавание символов.
12. Создание систем анализа и генерации текстов на основе искусственного интеллекта.
13. Оценка качества систем обработки и анализа текстов.
14. Этические аспекты использования искусственного интеллекта в анализе текстов.
15. Применение методов обработки естественного языка и машинного анализа текстов в реальных проектах.

## 4 Учебно-методическое и информационное обеспечение

### 4.1 Нормативные документы и ГОСТы

1. Федеральный закон от 29.12.2012 № 273-ФЗ «Об образовании в Российской Федерации».

2. Приказ Минобрнауки России от 19.09.2017 N 929 (ред. от 08.02.2021) «Об утверждении федерального государственного образовательного стандарта высшего образования – бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника» (Зарегистрировано в Минюсте России 10 октября 2017 г. N 48489).

3. Академический учебный план Направление подготовки: 09.03.01 Информатика и вычислительная техника Профиль: Системная и программная инженерия Форма обучения: очная.

4. Положение о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся федерального государственного автономного образовательного учреждения высшего образования «Московский политехнический университет» (Утверждено приказом Московского Политеха от 01.12.2022 № 1375ОД).

### 4.2 Основная литература


1. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/520544
2. Воронов, М. В. Системы искусственного интеллекта : учебник и практикум для вузов / М. В. Воронов, В. И. Пименов, И. А. Небаев. — Москва : Издательство Юрайт, 2023. — 256 с. — (Высшее образование). — ISBN 978-5-534-14916-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/519916

3. Иванов, В. М. Интеллектуальные системы : учебное пособие для вузов / В. М. Иванов ; под научной редакцией А. Н. Сесекина. — Москва : Издательство Юрайт, 2022. — 91 с. — (Высшее образование). — ISBN 978-5-534-00551-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/492094

### 4.3 Дополнительная литература

1. Бессмертный, И. А. Системы искусственного интеллекта : учебное пособие для вузов / И. А. Бессмертный. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 157 с. — (Высшее образование). — ISBN 978-5-534-07467-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/512657
2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/511020

### 4.4 Электронные образовательные ресурсы

Ссылка на электронный образовательный ресурс
https://lms.mospolytech.ru/enrol/index.php?id=14103

## 5 Материально-техническое обеспечение

### 5.1 Требования к оборудованию и помещению для занятий

Практические занятия (семинары) и самостоятельная работа студентов должна проводиться в специализированных аудиториях с комплектом мультимедийного оборудования и/или доской для записей материалов. Число рабочих мест в аудитории должно быть достаточным для обеспечения индивидуальной работы студентов.

### 5.2 Требования к программному обеспечению

Для проведения практических занятий (семинаров) специального программного обеспечения для освоения дисциплины не требуется.

## 6 Методические рекомендации

### 6.1 Методические рекомендации для преподавателя по организации обучения

1. При подготовке к занятиям следует предварительно проработать материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия. Следует подготовить необходимые материалы – теоретические сведения, задачи и др. При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.

2. При проверке работ и отчетов следует учитывать не только правильность выполнения заданий, но и оптимальность выбранных методов решения, правильность выполнения всех его шагов.

### 6.2 Методические указания для обучающихся по освоению дисциплины

Изучение дисциплины осуществляется в строгом соответствии с целевой установкой в тесной взаимосвязи учебным планом. Основой теоретической подготовки студентов являются аудиторные занятия, семинары и практики.

В процессе самостоятельной работы студенты закрепляют и углубляют знания, полученные во время аудиторных занятий, дорабатывают конспекты и записи, готовятся к промежуточной аттестации, а также самостоятельно изучают отдельные темы учебной программы.

На занятиях студентов, в том числе предполагающих практическую деятельность, осуществляется закрепление полученных, в том числе и в процессе самостоятельной работы, знаний. Особое внимание обращается на развитие умений и навыков установления связи положений теории с профессиональной деятельностью будущего специалиста.

Самостоятельная работа осуществляется индивидуально. Контроль самостоятельной работы организуется в двух формах:

- самоконтроль и самооценка студента;
- контроль со стороны преподавателей (текущий и промежуточный).

Текущий контроль осуществляется на аудиторных занятиях, промежуточный контроль осуществляется на зачете в письменной (устной) форме.

Критериями оценки результатов самостоятельной работы студента являются:

- уровень освоения студентом учебного материала;
- умения студента использовать теоретические знания при выполнении практических задач;
- сформированность компетенций;
- оформление материала в соответствии с требованиями.

## 7 Фонд оценочных средств

### 7.1 Методы контроля и оценивания результатов обучения

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

- курсовая работа;
- экзамен.

### 7.2 Шкала и критерии оценивания результатов обучения

Показателем оценивания компетенций на различных этапах их формирования является достижение обучающимися планируемых результатов обучения по дисциплине.

| **Показатель** | **Критерии оценивания** | | | |
|---|---|---|---|---|
| | **2** | **3** | **4** | **5** |
| ОПК-8. Способен разрабатывать алгоритмы и программы, пригодные для практического применения | | | | |

| | | | | |
|---|---|---|---|---|
| ИОПК-8.1. Знает основные технологии создания и внедрения информационных систем, стандарты управления жизненным циклом информационной системы.<br>ИОПК-8.2. Умеет проектировать блок-схемы алгоритмов, оценивать производительность алгоритмов и затраты памяти на работу алгоритма, разрабатывать программы на основе спроектированного алгоритма и проводить отладку программы, применять методы системного анализа и математического моделирования при разработке и эксплуатации ИС, проводить структурный анализ, функциональный анализ, объектно-ориентированный анализ иерархии классов. осуществлять организационное обеспечение выполнения работ на всех стадиях и в процессах жизненного цикла | Обучающийся демонстрирует полное отсутствие или недостаточное соответствие материалу дисциплины знаний, указанных в индикаторах компетенций дисциплины. | Обучающийся демонстрирует неполное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации. | Обучающийся демонстрирует частичное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины. Но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях. | Обучающийся демонстрирует полное соответствие следующих знаний, указанных в индикаторах компетенций дисциплины. Свободно оперирует приобретенными знаниями. |

| | | | | |
|---|---|---|---|---|
| информационной системы.<br>ИОПК-8.3. Владеет навыками разработки программ, построения блок-схем алгоритмов и оценки производительности алгоритмов, работы с унифицированным языком визуального моделирования, составления плановой и отчетной документации по управлению проектами создания информационных систем на стадиях жизненного цикла. | | | | |

Шкала оценивания результатов промежуточной аттестации и её описание:

**Форма промежуточной аттестации: экзамен.**

Промежуточная аттестация обучающихся в форме экзамена проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по данной дисциплине (модулю), при этом учитываются результаты текущего контроля успеваемости в течение семестра. Оценка степени достижения обучающимися планируемых результатов обучения по дисциплине (модулю) проводится преподавателем, ведущим занятия по дисциплине (модулю) методом экспертной оценки. По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно».

К промежуточной аттестации допускаются только студенты, выполнившие все виды учебной работы, предусмотренные рабочей программой по дисциплине.

| Шкала оценивания | Описание |
|---|---|
| Отлично | Среднее значение для всех формируемых на момент проведения аттестации уровней компетенций – 5. Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации. |

| | |
|---|---|
| Хорошо | Среднее значение для всех формируемых на момент проведения аттестации уровней компетенций – 4. Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует неполное, правильное соответствие знаний, умений, навыков приведенным в таблицах показателей, либо если при этом были допущены 2-3 несущественные ошибки. |
| Удовлетворительно | Среднее значение для всех формируемых на момент проведения аттестации уровней компетенций – 3. Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, в котором освещена основная, наиболее важная часть материала, но при этом допущена одна значительная ошибка или неточность. |
| Неудовлетворительно | Не достигнуто пороговое значение хотя бы для одного уровня формируемых на момент проведения аттестации компетенций. Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации. |

## 7.3 Оценочные средства

### 7.3.1 Экзаменационные вопросы

1. Что такое обработка естественного языка?
2. Какие основные задачи стоят перед машинным анализом текстов?
3. Что включает в себя распознавание языка?
4. В чем суть классификации текстов?
5. Как осуществляется кластеризация текстовых данных?
6. Что значит извлечение информации из текста?
7. Что понимается под рекомендательной системой на основе анализа текста?
8. Какие задачи решает автоматический перевод и аннотирование текста?
9. Что входит в информационный поиск с использованием методов обработки естественного языка?
10. Что представляют собой чат-боты и системы разговорных интерфейсов?
11. В чем заключается обработка текста на изображении?
12. Для чего нужны системы анализа и генерации текста на основе искусственного интеллекта?
13. Как оценить качество системы обработки и анализа текста?
14. Каковы этические аспекты применения искусственного интеллекта в анализе текста?
15. Какие существуют подходы к обработке естественного языка?
16. Какие алгоритмы машинного обучения применяются в обработке естественного языка и машинном анализе текстов?
17. Какие инструменты и библиотеки используются для обработки естественного языка на языке Python?

18. Что представляет собой классификация текстов с использованием машинного обучения?
19. Как работает кластеризация текста с использованием алгоритмов машинного обучения?
20. В каких сферах применяются системы обработки естественного языка и машинного анализа текста?
21. Какие функции выполняют рекомендательные системы на основе анализа текстовых данных?
22. Как происходит автоматический перевод текста с использованием нейронных сетей?
23. Какие подходы используются для создания систем генерации текста на основе машинного обучения?
24. Что означает понятие “обработка текста на изображении” в контексте обработки естественного языка?
25. В чём заключается работа чат-бота с использованием обработки естественного языка?
26. Как можно оценить качество работы системы обработки и анализа текстовой информации?
27. Каким образом происходит извлечение информации из текстовой информации с помощью машинного обучения?
28. В каких случаях применяются системы анализа и генерации текстов?
29. Какие технологии используются для создания чат-ботов с использованием обработки естественного языка?
30. Какие принципы и задачи лежат в основе машинного анализа текстов?

### 7.3.2 Типовые практические задания

1. Разработка алгоритма машинного обучения для классификации текстов по жанрам (например, роман, детектив, научная фантастика).
2. Создание системы автоматического аннотирования научных статей на основе анализа их заголовков и ключевых слов.
3. Реализация системы распознавания языка с использованием метода n-грамм и машинного обучения для определения языка текста.
4. Разработка рекомендательной системы для книг на основе анализа пользовательских предпочтений и отзывов других читателей.
5. Создание чат-бота для поддержки пользователей на основе обработки естественного языка с использованием правил и машинного обучения.
6. Разработка системы машинного анализа текстов для поиска информации в больших текстовых коллекциях.
7. Создание системы машинного перевода текста на основе использования нейросетей и больших языковых моделей.
8. Реализация системы генерации текстов на заданную тему с использованием шаблонов и машинного обучения (например, генерация описаний товаров для интернет-магазина).
9. Разработка программы для анализа тональности текстов с использованием методов машинного обучения и обработки естественного языка (например, определение позитивного, негативного или нейтрального тона текста).
10. Создание системы анализа и классификации текстов на основе алгоритмов машинного обучения и лингвистических методов.

### 7.3.3 Типовой экзаменационный билет

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
**«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**
**(МОСКОВСКИЙ ПОЛИТЕХ)**

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1
по дисциплине
«Обработка естественного языка и машинный анализ текстов»
направление подготовки 09.03.01 Информатика и вычислительная техника

ВОПРОСЫ:

1. Какие принципы и задачи лежат в основе машинного анализа текстов?
2. Как происходит автоматический перевод текста с использованием нейронных сетей?
3. Что значит извлечение информации из текста?

Утверждено: _____ / _____ / «__» _____ 20__г.