

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики**

А.М. Райгородский

	Рабочая программа дисциплины (модуля)
по дисциплине:	Алгоритмы обработки строк и теории чисел
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Системное программирование и прикладная математика Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования
курс:	2
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 4 (весенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 30 час.

Самостоятельная работа: 90 час.

Подготовка к экзамену: 30 час.

Всего часов: 180, всего зач. ед.: 4

Количество контрольных работ, заданий: 3

Программу составил: С.Л. Бабичев, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры алгоритмов и технологий программирования 07.04.2022

Аннотация

Курс представляет собой вторую часть большого курса по разработке и анализу алгоритмов. Она существенно опирается на аппарат, полученный студентами на курсе «Алгоритмы на графах и динамическое программирование».

Семестр посвящен изучению классических алгоритмов и их реализаций, которые применимы к строкам и похожим на них структурам данных (цепочки сравнимых объектов).

Как и в предыдущих семестрах, для всех алгоритмов подробно доказываются корректность и асимптотики.

1. Цели и задачи

Цель дисциплины

Целями дисциплины являются первичное ознакомление студентов с основными принципами проектирования и анализа алгоритмов и структур данных, обучение навыкам обоснования корректности алгоритмов, их практической реализации, теоретической и экспериментальной оценки их временной сложности.

Задачи дисциплины

- научить формулировать задачи в терминах изученных теорий, выбирать подходящий алгоритм для поставленной задачи;
- изучить такие разделы, как хеш-таблицы, динамическое программирование и графы;
- изучить такие разделы, как прикладные задачи теории чисел и комбинаторики, поисковые задачи на строках, сжатие текста, вычислительная геометрия и эвристические алгоритмы.
- научить разрабатывать композиции алгоритмов для решения поставленных задач, оценивать сложности алгоритмов, выбирать подходящие структуры данных для поставленных задач, реализовывать алгоритмы на языках программирования C и C++.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1 Формулирует совокупность взаимосвязанных задач в рамках поставленной цели работы, обеспечивающих ее достижение. Определяет ожидаемые результаты решения поставленных задач
	УК-2.2 Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.2 Знает основные источники научно-технической и (или) технологической информации в области профессиональной деятельности

ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
---	---

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- определения асимптотик в O-нотации;
- определения простейших линейных структур данных (стек, очередь, дек, вектор) и сложность обработки запросов в них;
- алгоритм быстрой сортировки;
- алгоритм сортировки слиянием
- определение и практическую необходимость деревьев поиска;

- классификацию и сферы применения хеш-функций;
- организацию и использование хеш-таблиц;
- условия возникновения задачи динамического программирования;
- сведение задачи динамического программирования к функции Беллмана;
- виды задач динамического программирования и методы их решения;
- понятие графов и организацию их хранения;
- алгоритмы обхода графов;
- алгоритмы нахождения специальных элементов в графах;
- алгоритмы нахождения минимальных остовных деревьев;
- алгоритмы нахождения кратчайших путей в графах;
- алгоритмы нахождения потоков в графах;
- алгоритмы нахождения минимальных разрезов и максимальных потоков в графах.
- строго доказывать утверждения о корректности алгоритмов;
- применять необходимую технику для решения алгоритмических задач;

- основы теории чисел в применении к информационным технологиям;
- быстрые алгоритмы работы с длинными числами и матрицами;
- применение алгоритмов теории чисел к задачам криптографии;
- методы комбинаторного поиска;
- Z-функцию и префикс-функцию и их применение;
- структуру данных бор и алгоритм Ахо-Корасик;
- суффиксное дерево и суффиксный автомат;
- алгоритмы сжатия информации: Huffman, LZ77, LZ78, BWT, Arithmetic coding;
- алгоритмы вычислительной геометрии: нахождение выпуклой оболочки, триангуляцию, поиск на плоскости.
- методы приближённого решения NP-сложных задач

уметь:

- оценивать сложность алгоритмов;
- строго доказывать утверждения о корректности алгоритмов;
- применять необходимую технику для решения алгоритмических задач;

владеть:

- разнообразными деревьями поиска и методикой выбора наиболее предпочтительного в каждой конкретной ситуации;
 - методами доказательства корректности утверждений об алгоритмах;
 - методами декомпозиции задач на более простые;
 - методами синтеза решения сложных задач из простых.
-
- техникой выбора необходимой хеш-функции для данной задачи;
 - техникой сведения уравнения Беллмана к программному коду;
 - техникой реализаций изученных алгоритмов на графах;
 - методами доказательства корректности утверждений об алгоритмах;
 - методами декомпозиции задач на более простые;
 - методами синтеза решения сложных задач из простых.
-
- техникой выбора необходимых алгоритмов для решения задач теории чисел;
 - техникой быстрого поиска информации в текстовых данных;
 - умением обратимо преобразовывать информацию в более компактное представление;
 - методами аналитической геометрии для решения задач вычислительной геометрии;
 - умением находить приближённые решения сложных задач эвристическими методами.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Теория чисел.	5		5	15
2	Строки. Z-функция.	5		5	15
3	Структура данных бор.	5		5	15
4	Алгоритмы сжатия текста.	5		5	15
5	Вычислительная геометрия.	5		5	15
6	Приближённое решение неполиномиальных алгоритмов.	5		5	15
Итого часов		30		30	90
Подготовка к экзамену		30 час.			
Общая трудоёмкость		180 час., 4 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 4 (Весенний)

1. Теория чисел.

Модульная арифметика. Быстрое возведение в степень. Схема Горнера. Решето Эратосфена. Алгоритм Евклида. Диофантовы уравнения. Побитовое нахождение НОД. Порядок элемента по модулю. Определение простоты числа. Факторизация. Алгоритмы арифметики длинных чисел. Нахождение первообразного корня. Прямое и обратное дискретное преобразование Фурье. Быстрое умножение многочленов. Быстрое умножение матриц. Алгоритм Штрассена. Комбинаторный поиск. Генерирование всех перестановок. Алгоритм Диффи-Хеллмана. Алгоритм RSA.

2. Строки. Z-функция.

Префикс-функция. Преобразование Z-функции в префикс-функцию и обратно. Алгоритм Кнута-Мориса-Пратта. Rolling-hash. Алгоритм Карпа-Рабина.

3. Структура данных бор.

Алгоритм Ахо-Корасик. Сопоставление строк по образцу. Вхождение образца в строку. Подсчёт числа вхождений. Суффиксное дерево и суффиксный автомат.

4. Алгоритмы сжатия текста.

Алгоритм Хаффмена. Алгоритм LZ78. Алгоритм LZ77 с вариациями (LZSS). Преобразование Burrows-Wheeler. Арифметическое кодирование. Контекстное кодирование.

5. Вычислительная геометрия.

Представление геометрических примитивов. Скалярное и векторное произведения. Пересечение двух прямых. Пересечение прямой и окружности. Пересечение двух окружностей. Пересечение двух отрезков. Площади многоугольников. Метод шнуровки. Нахождение принадлежности точки выпуклому многоугольнику. Нахождение принадлежности точки произвольному многоугольнику. Метод пересекающей прямой. Выпуклая оболочка. Алгоритмы Джарвиса, Грэхема, Эндрю. Триангуляция выпуклой оболочкой. Нахождение диаметра множества точек. Метод вращающихся калиперов. Нахождение кратчайшего расстояния между точками множества: жадный алгоритм и алгоритм разделяй и властвуй.

6. Приближённое решение неполиномиальных алгоритмов.

Перебор с отсечением. Альфа-бета алгоритм. Задача о вершинном покрытии графа. Имитация отжига. Обратные задачи. Генетические алгоритмы. Метод роя частиц. Метод дифференциальной эволюции.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория с доской.

6. Перечень рекомендуемой литературы

Основная литература

1. Алгоритмы : построение и анализ [Текст] / Т. Кормен, Ч. Лейзерсон, Р. Ривест .— М. : МЦНМО, 2001 .— 960 с.
2. Программирование: теоремы и задачи [Текст], [учеб. пособие] /А. Шень. -М., МЦНМО, 2017
3. Алгоритмы [Текст] : [учеб. пособие для вузов] / С. Дасгупта, Х. Пападимитриу, У. Вазиани ; пер. с англ. А. А. Куликова ; под ред. А. Шеня .— М. : МЦНМО, 2014 .— 320 с.

Дополнительная литература

Не предусмотрено

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. http://neerc.ifmo.ru/wiki/index.php?title=Дискретная_математика,_алгоритмы_и_структуры_данных. «Викиконспекты», сайт Санкт-Петербургского Университета ИМТО.
2. <http://e-maxx.ru/>. Maximal algo: личный сайт Максима Иванова, посвящённый алгоритмам и структурам данных.

3. <https://codeforces.com/>. Международная платформа онлайн-кон테стов по спортивному программированию на базе СГУ и ИМТО.

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

Для успешного освоения дисциплины, студент использует следующие программные средства: компилятор языка С.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

В преподавании дисциплины сочетаются практические занятия в компьютерном классе и домашние задания, состоящие из теоретических задач и задач на программирование с автоматической проверкой корректности.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Информатика и вычислительная техника
профиль подготовки:	Системное программирование и прикладная математика Физтех-школа Прикладной Математики и Информатики кафедра алгоритмов и технологий программирования
курс:	2
квалификация:	бакалавр

Семестр, формы промежуточной аттестации: 4 (весенний) - Экзамен

Разработчик: С.Л. Бабичев, канд. физ.-мат. наук, доцент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1 Формулирует совокупность взаимосвязанных задач в рамках поставленной цели работы, обеспечивающих ее достижение. Определяет ожидаемые результаты решения поставленных задач
	УК-2.2 Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.2 Знает основные источники научно-технической и (или) технологической информации в области профессиональной деятельности
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации

2. Показатели оценивания компетенций

В результате изучения дисциплины «Алгоритмы обработки строк и теории чисел» обучающийся должен:

знать:

- определения асимптотик в О-нотации;
- определения простейших линейных структур данных (стек, очередь, дек, вектор) и сложность обработки запросов в них;
- алгоритм быстрой сортировки;
- алгоритм сортировки слиянием
- определение и практическую необходимость деревьев поиска;
- классификацию и сферы применения хеш-функций;
- организацию и использование хеш-таблиц;
- условия возникновения задачи динамического программирования;
- сведение задачи динамического программирования к функции Беллмана;
- виды задач динамического программирования и методы их решения;
- понятие графов и организацию их хранения;
- алгоритмы обхода графов;
- алгоритмы нахождения специальных элементов в графах;
- алгоритмы нахождения минимальных остовных деревьев;
- алгоритмы нахождения кратчайших путей в графах;
- алгоритмы нахождения потоков в графах;
- алгоритмы нахождения минимальных разрезов и максимальных потоков в графах.
- строго доказывать утверждения о корректности алгоритмов;
- применять необходимую технику для решения алгоритмических задач;
- основы теории чисел в применении к информационным технологиям;
- быстрые алгоритмы работы с длинными числами и матрицами;
- применение алгоритмов теории чисел к задачам криптографии;
- методы комбинаторного поиска;
- Z-функцию и префикс-функцию и их применение;
- структуру данных бор и алгоритм Ахо-Корасик;
- суффиксное дерево и суффиксный автомат;
- алгоритмы сжатия информации: Huffman, LZ77, LZ78, BWT, Arithmetic coding;
- алгоритмы вычислительной геометрии: нахождение выпуклой оболочки, триангуляцию, поиск на плоскости.
- методы приближённого решения NP-сложных задач

уметь:

- оценивать сложность алгоритмов;
- строго доказывать утверждения о корректности алгоритмов;
- применять необходимую технику для решения алгоритмических задач;

владеть:

- разнообразными деревьями поиска и методикой выбора наиболее предпочтительного в каждой конкретной ситуации;
- методами доказательства корректности утверждений об алгоритмах;
- методами декомпозиции задач на более простые;
- методами синтеза решения сложных задач из простых.
- техникой выбора необходимой хеш-функции для данной задачи;
- техникой сведения уравнения Беллмана к программному коду;
- техникой реализаций изученных алгоритмов на графах;
- методами доказательства корректности утверждений об алгоритмах;
- методами декомпозиции задач на более простые;
- методами синтеза решения сложных задач из простых.
- техникой выбора необходимых алгоритмов для решения задач теории чисел;
- техникой быстрого поиска информации в текстовых данных;
- умением обратимо преобразовывать информацию в более компактное представление;
- методами аналитической геометрии для решения задач вычислительной геометрии;
- умением находить приближённые решения сложных задач эвристическими методами.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

для 4-го семестра

1. Найдите произведение данных многочленов [многочлены] с использованием дискретного преобразования Фурье.
2. Определите последовательность токенов в алгоритме LZ77 для следующего текста: [дан фрагмент текста]
3. Приведите пример множества точек, для которого жадный алгоритм нахождения минимального расстояния между точками сходится медленно.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

примеры экзаменационных вопросов для 4-го семестра

- техникой выбора необходимых алгоритмов для решения задач теории чисел;
- техникой быстрого поиска информации в текстовых данных;
- умением обратимо преобразовывать информацию в более компактное представление;
- методами аналитической геометрии для решения задач вычислительной геометрии;
- умением находить приближённые решения сложных задач эвристическими методами.

1. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

1. Найдите произведение данных многочленов [многочлены] с использованием дискретного преобразования Фурье.
2. Определите последовательность токенов в алгоритме LZ77 для следующего текста: [дан фрагмент текста]
3. Приведите пример множества точек, для которого жадный алгоритм нахождения минимального расстояния между точками сходится медленно.

2. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Построить Z-функцию для данной строки.
2. На столе находятся блины - круги с заданными центром и радиусом. Определите, сколько блинов будет разрезано лучом, проведённым из заданной точки в другую заданную.
3. На плоскости заданы множество P точек. Определить для заданных точек ближайшие к ним из множества P . Установите сложность подсчёта, если он нужен, и сложность поиска.

Пример экзаменационных вопросов.

1. Алгоритм побитового нахождения НОД. Доказательство корректности и определение сложности.
2. Дискретное преобразование Фурье. Алгоритмы и применение.
3. Алгоритмы деления длинных чисел.
4. Алгоритм Штрассена. Доказательство его корректности и определение сложности.
5. Вычисление функции Эйлера для заданного числа.
6. Алгоритм Диффи-Хеллмана и его применения.
7. Алгоритм RSA и его применения.
8. Z-функция. Свойства и вычисление за $O(N)$. Применение.
9. Префикс-функция. Свойства и вычисление за $O(N)$. Применение. Алгоритм Кнута-Мориса-Пратта.
10. Структура данных бор. Алгоритм Ахо-Корасик. Применение.
11. Суффиксное дерево и суффиксный автомат.
12. Алгоритм Хаффмена. Кодирование таблицы частот.
13. Алгоритм LZ78. Варианты алгоритма. Борьба с переполнением.
14. Алгоритм LZ77. Алгоритмы быстрого поиска подстрок. Варианты.
15. Арифметическое кодирование. Применение в контекстном сжатии.

16. Вычисление площади ориентированного многоугольника. Формула Пика. Алгоритм шнуровки.
17. Определение принадлежности точки многоугольнику: выпуклому и невыпуклому.
18. Нахождение выпуклой оболочки. Алгоритм Джарвиса.
19. Нахождение выпуклой оболочки. Алгоритм Грэхема.
20. Алгоритм триангуляции множества точек на плоскости.
21. Определение диаметра множества точек на плоскости.
22. Нахождение самых близких точек на плоскости: два варианта поиска.
23. Перебор с отсечением. Альфа-бета алгоритм.
24. Приближённое решение NP-сложных. Имитация отжига.
25. Приближённое решение NP-сложных задач. Генетические алгоритмы.

пример экзаменационного билета

1. Построить Z-функцию для данной строки.
2. На столе находятся блины - круги с заданными центром и радиусом. Определите, сколько блинов будет разрезано лучом, проведённым из заданной точки в другую заданную.
3. На плоскости заданы множество P точек. Определить для заданных точек ближайшие к ним из множества P. Установите сложность предподсчёта, если он нужен, и сложность поиска.

Критерии оценивания

отлично

- 10 Полностью и вовремя решены все задачи без ошибок. Продемонстрирован грамотный подход к решению задач, реализованы оптимальные алгоритмы, код оформлен в едином удобочитаемом стиле
- 9 Полностью и вовремя решены все задачи без ошибок. Продемонстрирован грамотный подход к решению задач, реализованы оптимальные алгоритмы
- 8 Полностью и вовремя решены все задачи без ошибок. Продемонстрирован грамотный подход к решению задач

хорошо

- 7 Полностью решены все задачи. Допущены несущественные ошибки.
- 6 Полностью решено большинство задач. В некоторых задачах допущены и не исправлены ошибки, либо некоторые задачи решены частично.
- 5 Полностью решено две трети задач. В некоторых задачах допущены и не исправлены ошибки, либо некоторые задачи решены частично.

удовлетворительно

- 4 Полностью решено более половины задач. В остальных задачах допущены и не исправлены ошибки, либо некоторые задачи решены частично.
- 3 Полностью решено более половины задач.

неудовлетворительно

- 2 Решено менее половины задач.
- 1 Не решено ни одной задачи.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

При проведении устного экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на устном экзамене не должен превышать двух астрономических часов.

Во время проведения экзамена обучающиеся могут пользоваться программой дисциплины.