

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Институт информационных технологий, математики и механики

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Строковые алгоритмы

Уровень высшего образования

Магистратура

Направление подготовки / специальность

01.04.02 - Прикладная математика и информатика

Направленность образовательной программы

Анализ данных в прикладных областях

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.02.03 Строковые алгоритмы относится к части, формируемой участниками образовательных отношений образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ПК-12: Способен разрабатывать и применять математические методы, системное и прикладное программное обеспечение для решения задач производственно-технологической деятельности	<p>ПК-12.1: Знает основные методы разработки математических методов, системного и прикладного программного обеспечения для решения задач производственно-технологической деятельности</p> <p>ПК-12.2: Умеет оценивать трудоемкость разработки программных средств для решения задач производственно-технологической деятельности</p> <p>ПК-12.3: Имеет навыки разработки системного программного обеспечения для решения задач производственно-технологической деятельности</p>	<p>ПК-12.1: ЗНАТЬ основные методы разработки и анализа алгоритмических, методических и технологических проблем</p> <p>ПК-12.2: УМЕТЬ оценивать трудоемкости разработки программных средств для решения задач научной и производственно-технологической деятельности</p> <p>ПК-12.3: Владеет навыками разработки концептуальных и теоретических моделей решаемых проблем и задач алгоритмов на строках; владеет опытом реализации программных систем для решения практических задач теории строковых алгоритмов</p>	Задачи	Зачёт: Контрольные вопросы
ПК-5: Способен разрабатывать и применять математические	ПК-5.1: Знает типовые математические методы и методологии разработки	ПК-5.1: ЗНАТЬ базовые структуры данных и	Задачи	Зачёт: Контрольные

методы, системное и прикладное программное обеспечение для решения задач научной деятельности	системного и прикладного программного обеспечения для решения задач научной деятельности ПК-5.2: Умеет применять типовые математические методы и методологии разработки системного и прикладного программного обеспечения для решения задач научной деятельности ПК-5.3: Имеет навыки разработки системного и прикладного программного обеспечения для решения задач научной деятельности	алгоритмы для работы с информацией, представленной в виде последовательностей символов ПК-5.2: УМЕТЬ профессионально разрабатывать и использовать программное обеспечение для решения задач научной и проектно-технологической деятельности ПК-5.3: Владеет методами применения техник хеширования и техник дискретного преобразования Фурье для решения задач на строках; постановки задач нечеткого сопоставления на строках		вопросы
---	---	--	--	---------

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	4
Часов по учебному плану	144
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	32
- занятия семинарского типа (практические занятия / лабораторные работы)	32
- КСР	1
самостоятельная работа	79
Промежуточная аттестация	0 Зачёт

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего	в том числе
--	-------	-------------

	(часы)	Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Алгоритмы и структуры данных поиска	35	8	8	16	19
Алгоритмы для работы с большими данными	36	8	8	16	20
Графовые алгоритмы	36	8	8	16	20
Строковые алгоритмы	36	8	8	16	20
Аттестация	0				
КСР	1			1	
Итого	144	32	32	65	79

Содержание разделов и тем дисциплины

Период строки. Префикс функция. Z-функция. Trie(бор) структура. Алгоритм Ахо-Корасик. Функция ошибок. Суффиксный бор. Алгоритм Укконена. Задача о наибольшей общей подстроке. RMQ, SMQ и LCA структуры. Статический и динамический варианты. Суффиксный массив. Построение суффиксного массива за линейное время. Построение LCP массива за линейное время. Применение техник хеширования для строковых задач. Алгоритм Рабина-Карпа. Приближенное сопоставление строк с джокерами. Разные алгоритмы. Приближенное сопоставление строк по метрикам Хемминга и Левенштейна. Задача о наибольшей общей подпоследовательности. Алгоритм Ханта-Сцимански, метод четырех русских. Построение сжатых индексов. Преобразование Барроуза-Уилера. Wavelet tree. Конечные автоматы, введение. Минимальный автомат. Алгоритм Мура, Алгоритм Хопкрофта. Недетерминированные конечные автоматы. Регулярные выражения и регулярные языки. Сведение регулярного выражения к НКА. Сведение ДКА к регулярному выражению. Лемма о накачке. Контекстно свободные грамматики. Сведение к нормальной форме Хомского. Алгоритм СЮК.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

При выполнении самостоятельной работы студентам рекомендуется использовать курс видеолекций <https://www.youtube.com/playlist?list=PLCs9YfffJvSRfmICgeymMgnD7ubOkjBAz>

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Задачи) для оценки сформированности компетенции ПК-12:

1. Построение суффиксного дерева по набору текстов.
2. Моделирование прохода по образцу.

5.1.2 Типовые задания (оценочное средство - Задачи) для оценки сформированности компетенции ПК-5:

1. Построение префиксного бора Ахо-Корасик.
2. Моделирование прохода Ахо-Корасик.

Критерии оценивания (оценочное средство - Задачи)

Оценка	Критерии оценивания
зачтено	Задача решена полностью или решена основная часть задачи, или задача решена с недочетами
не зачтено	Задача не решена или сделан первый этап решения задачи

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие	При решении стандартных задач не продемонстрированы основные умения. Имели	Продемонстрированы основные умения. Решены типовые задачи с	Продемонстрированы все основные умения. Решены все основные	Продемонстрированы все основные умения. Решены все основные	Продемонстрированы все основные умения. Решены все	Продемонстрированы все основные умения. Решены все основные задачи.

	отказа обучающегося от ответа	место грубые ошибки	негрубыми ошибками. Выполнены все задания, но не в полном объеме	задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	задачи. Выполнены все задания в полном объеме, но некоторые с недочетами	основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован творческий подход к решению нестандартных задач

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-12

1. Сформулируйте понятие периода и бордера строки. Сформулируйте понятия префикс функции и Z-функции. Приведите линейные алгоритмы построения последних.
2. Опишите алгоритм построения префиксного бора по набору образцов. Сформулируйте понятие функции ошибок. Опишите алгоритм Ахо-Корасик.
3. Дайте определение суффиксного бора. Каким образом можно осуществить сжатие суффиксного бора до линейной памяти? Приведите алгоритм Укконена.
4. Дайте определение суффиксного массива. Каким образом и за какое время можно решать задачу о подстроке пользуясь суффиксным массивом? Расскажите как строить суффиксный массив за линейное время.
5. Дайте определения SMQ, RMQ и LCA структур, их статических и динамических вариантов. Какую сложность имеют основные операции данных структур в статическом и динамическом вариантах? Приведите конкретные методы приводящие к этим трудоемкостям.

5.3.2 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-5

1. Дайте определение LCP массива. Приведите линейный алгоритм построения LCP массива. Для каких задач используется LCP-массив?
2. Сформулируйте задачу сопоставления двух строк с джокерами. Приведите алгоритм основанный на алгоритме Ахо-Корасик. Приведите алгоритм основанный на быстром преобразовании Фурье. Приведите алгоритм основанный на технике полиномиального хеширования.
3. Сформулируйте задачу сопоставления двух строк с не более чем k ошибками. Приведите алгоритм который решает данную задачу за время $O(nk)$ с использованием LCP-структуры для метрики Хемминга. Приведите алгоритм Ландау-Вишкина, который решает аналогичную задачу для метрики Левенштейна.
4. Расскажите о том как техники хеширования могут применяться для задач на строках. Расскажите о полиномиальном хешировании. Расскажите алгоритм Рабина-Карпа.
5. Опишите задачу построения сжатого индекса. Каким образом это осуществимо? Расскажите о преобразовании Барроуза-Уилера. Опишите структуру данных wavelet tree.
6. Сформулируйте задачи о наибольшей общей подстроке и о наибольшей общей подпоследовательности. Как первая из этих задач может быть решена за линейное время. Приведите алгоритм Ханта-Сцимански и метод «четырёх русских» для второй задачи.
7. Дайте определения детерминированного и недетерминированного конечных автоматов. Дайте определения регулярных выражений и языков. Дайте определение минимального автомата. Приведите алгоритмы минимизации Мура и Хопкрофта.
8. Опишите метод сведения регулярного выражения к недетерминированному конечному автомату. Опишите метод сведения регулярного выражения к детерминированному конечному автомату. Сформулируйте и докажите лемму о накачке.

9. Дайте определения и примеры контекстно свободных грамматик. Дайте определение нормальной формы Хомского. Как осуществляется приведение к нормальной форме Хомского. Расскажите алгоритм СΥК разбора грамматики.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Владение основным и дополнительным материалом достаточное или с незначительными ошибками и погрешностями
не зачтено	владение материалом, необходимым по данному предмету, недостаточно. Работу за время семестра можно оценить как неудовлетворительную

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Алексеев Владимир Евгеньевич. Графы. Модели вычислений. Структуры данных : учеб. для студентов, обучающихся по специальности 010200- Приклад. математика и 510200-Приклад. математика и информатика / Нижегородский государственный университет им. Н. И. Лобачевского. - Н. Новгород : Изд-во ННГУ, 2005. - 307 с. : ил. - ISBN 5-85747-810-8 : 100.00., 2 экз.

Дополнительная литература:

1. Кнут Д. Э. Искусство программирования : учеб. пособие : пер. с англ. Т. 3. Сортировка и поиск. - 2-е изд., испр. и доп. - М. : Вильямс, 2000. - 832 с. - ISBN 0-201-89685-0 (англ.) : 317.00., 1 экз.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. <http://logic.pdmi.ras.ru/~yura/strings.html>
2. <https://www.coursera.org/learn/algorithms-on-strings>

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 01.04.02 - Прикладная математика и информатика.

Автор(ы): Грибанов Дмитрий Владимирович, кандидат физико-математических наук.

Заведующий кафедрой: Золотых Николай Юрьевич, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 13.12.2023, протокол № 3.