

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 Ученого совета ННГУ
протокол № 10 от 02.12.2024 г.

Рабочая программа дисциплины

Строковые алгоритмы

Уровень высшего образования
Магистратура

Направление подготовки / специальность
02.04.02 - Фундаментальная информатика и информационные технологии

Направленность образовательной программы
Искусственный интеллект

Форма обучения
очная

г. Нижний Новгород

2025 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.08.02 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-8: Способен к разработке новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности	<p>ПК-8.1: Знает методику разработки новых алгоритмических, методических и технологических решений</p> <p>ПК-8.2: Умеет применять полученные знания для разработки новых алгоритмических, методических и технологических решений</p> <p>ПК-8.3: Имеет практический опыт составления технического задания на разработку информационной системы</p>	<p>ПК-8.1:</p> <p>Знать основные алгоритмы поиска строки эквивалентной образцу в тексте: алгоритм КМП, алгоритм Ахо-Корасик; основные алгоритмы и структуры данных для решения задачи о подстроке: суффиксное дерево, алгоритм Укконена, суффиксный массив, алгоритм Каркайна-Сандерса, алгоритм КарнаМиллера-Розенберга, LCP-структура, LCP-массив, алгоритм Аримур-Арикавы-Ли-Парка; алгоритмы и структуры данных для задач RMQ (range minimum query) и RSQ (range sum query); их динамические и статические варианты, применение для создания LCP-структур; Уметь разрабатывать и анализировать концептуальные и теоретические модели для строковых задач и алгоритмов; распознавать и анализировать сложность прикладных и теоретических задач теории строковых алгоритмов.</p> <p>ПК-8.2:</p> <p>Знать как применять преобразования Барроуза-</p>	Задачи	Экзамен: Контрольные вопросы

		<p>Уилера для построения сжатого индекса, алгоритмы быстрого вычисления этого преобразования; применение структуры данных Wavelet tree для построения сжатого индекса; определение регулярного языка, детерминированного и недетерминированного конечного автомата;</p> <p>Доказательство эквивалентности языков, задаваемых регулярными выражениями, недетерминированными и детерминированными конечными автоматами;</p> <p>алгоритмы минимизации детерминированных конечных автоматов: алгоритм Мура, алгоритм Хопкрофта;</p> <p>определение контекстносвободной грамматики; сведение к нормальной форме Хомского; алгоритм СЮК; Уметь использовать современные математические методы для анализа и разработки строковых алгоритмов.</p> <p>Владеть навыками работы с современными библиотеками и наборами прикладных программ для строковых задач.</p> <p>ПК-8.3:</p> <p>Знать методы применения техник хеширования и техник дискретного преобразования Фурье для решения задач на строках; постановки задач нечеткого сопоставления на строках: редакционное расстояние, расстояние Хэмминга, различные модернизации редакционного расстояния, задача с джокерами (wildcards); основные алгоритмы решения задачи нечеткого сопоставления: алгоритм</p>		
--	--	--	--	--

		<p>Ландау-Вишкина, LCP-структура, применение техники дискретного преобразования Фурье; основные алгоритмы быстрого подсчета редакционного расстояния и решения задачи о длиннейшей общей подпоследовательности двух строк: алгоритм Ханта-Сцимански, алгоритм с применением метода четырех русских; алгоритмы для решения задачи о длиннейшей общей подстроке: динамическое программирование, применение суффиксного дерева; постановку задачи построения сжатого индекса; Владеть способностью разрабатывать и анализировать концептуальные и теоретические модели решаемых проблем и задач алгоритмов на строках; владение опытом реализации программных систем для решения практических задач теории строковых алгоритмов.</p>		
--	--	---	--	--

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)	16
- КСР	2
самостоятельная работа	58
Промежуточная аттестация	36
	Экзамен

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 (часы)	в том числе			
		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
Период строки. Z/P-функции.	7	2	1	3	4
Бор (trie) структура. Алгоритм Ахо-Корасик. Функция ошибок.	7	2	1	3	4
Суффиксный бор. Алгоритм Укконена. Задача о наибольшей общей подстроке.	7	2	1	3	4
RMQ, SMQ и LCA структуры. Статический и динамический варианты.	7	2	1	3	4
Суффиксный массив.	7	2	1	3	4
Техники хеширования для строковых задач. Алгоритм Кнута-Морриса-Пратта. Алгоритм Рабина-Карпа.	7	2	1	3	4
Приближенное сопоставление строк с джокерами.	7	2	1	3	4
Приближенное сопоставление строк по метрикам Хемминга и Левенштейна.	7	2	1	3	4
Задача о наибольшей общей подпоследовательности. Алгоритм Ханта-Сцимански, метод четырех русских.	7	2	1	3	4
Построение сжатых индексов. Преобразование Барроуза-Уилера. Wavelet tree	7	2	1	3	4
Конечные автоматы, введение. Минимальный автомат. Алгоритм Мура, Алгоритм Хопкрофта.	8	3	1	4	4
Недетерминированные конечные автоматы. Регулярные выражения и регулярные языки.	8	3	1	4	4
Сведение регулярного выражения к НКА. Сведение ДКА к регулярному выражению. Лемма о накачке.	10	3	2	5	5
Контекстно свободные грамматики. Сведение к нормальной форме Хомского. Алгоритм СУК.	10	3	2	5	5
Аттестация	36				
КСР	2			2	
Итого	144	32	16	50	58

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 Кока-Янгера-Касами.

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-8:

1. Построение суффиксного дерева по набору текстов.
2. Моделирование прохода по образцу.
3. Построение префиксного бора Ахо-Корасик.
4. Моделирование прохода Ахо-Корасик.

Критерии оценивания (оценочное средство - Задачи)

Оценка	Критерии оценивания
зачтено	Работа выполнена в полном объеме и в срок, результаты работы алгоритма корректные на тестовых примерах, результаты работы представлены преподавателю.
не зачтено	Работа не выполнена или выполнена не в полном объеме (программа работает некорректно на тестовых примерах, результаты работы не представлены преподавателю).

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 отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения. Решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи. 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 некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продemonстрирован 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-8

1. 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.
2. Расскажите о том, как техники хеширования могут применяться для задач на строках. Расскажите о полиномиальном хешировании. Расскажите алгоритм Рабина-Карпа.
3. Опишите задачу построения сжатого индекса. Каким образом это осуществимо? Расскажите о преобразовании Барроуза-Уилера. Опишите структуру данных wavelet tree.
4. 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.
5. Дайте определения детерминированного и не детерминированного конечных автоматов. Дайте определения регулярных выражений и языков. Дайте определение минимального автомата. Приведите алгоритмы минимизации Мура и Хопкрофта.
6. Опишите метод сведения регулярного выражения к не детерминированному конечному автомату. Опишите метод сведения регулярного выражения к детерминированному конечному автомату. Сформулируйте и докажите лемму о накачке.
7. Дайте определения и примеры контекстно свободных грамматик. Дайте определение нормальной формы Хомского. Как осуществляется приведение к нормальной форме Хомского. Расскажите алгоритм СΥΚ разбора грамматики.

8. Сформулируйте понятие периода и бордера строки. Сформулируйте понятия префикс функции и Z-функции. Приведите линейные алгоритмы построения последних.
9. Опишите алгоритм построения префиксного бора по набору образцов. Сформулируйте понятие функции ошибок. Опишите алгоритм Ахо-Корасик.
10. Дайте определение суффиксного бора. Каким образом можно осуществить сжатие суффиксного бора до линейной памяти? Приведите алгоритм Укконена.
11. Дайте определение суффиксного массива. Каким образом и за какое время можно решать задачу о подстроке пользуясь суффиксным массивом? Расскажите, как построить суффиксный массив за линейное время.
12. 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.
13. Дайте определение LCP массива. Приведите линейный алгоритм построения LCP массива. Для каких задач используется LCP-массив?
14. 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.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
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».
неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».

Оценка	Критерии оценивания
плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

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, компьютерами.

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 02.04.02 - Фундаментальная информатика и информационные технологии.

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

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

Программа одобрена на заседании методической комиссии от 02.12.2024, протокол № 5.