

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 от  
«30» ноября 2022 г. № 13

**Рабочая программа дисциплины**

**Строковые алгоритмы**

---

Уровень высшего образования  
магистратура

---

Направление подготовки  
02.04.02 Фундаментальная информатика и информационные технологии

---

Направленность образовательной программы  
**Искусственный интеллект**

---

Форма обучения  
очная

---

Нижний Новгород  
2023 год

## 1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.07.01 «Строковые алгоритмы» относится к дисциплинам по выбору Блока 1, формируемой участниками образовательных отношений, «Дисциплины (модули)» направления подготовки 02.04.02 «Фундаментальная информатика и информационные технологии», направленность «Искусственный интеллект». Дисциплина преподается в 2 семестре.

№ Варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
1	Блок 1. Дисциплины (модули) Часть, формируемая участниками образовательных отношений. Дисциплина по выбору	Дисциплина Б1.В.ДВ.07.01 «Строковые алгоритмы» относится к части ООП направления подготовки 02.04.02 Фундаментальная информатика и информационные технологии, формируемой участниками образовательных отношений.

## 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	
ПК-8. Способен к разработке новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности.	ПК-8.1. Знать основы разработки новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности	<u>Знать</u> 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 Каркайнена-Сандерса, алгоритм Карпа-Миллера-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; <u>Уметь</u> разрабатывать и анализировать концептуальные и теоретические модели для строковых задач и алгоритмов; распознавать и анализировать сложность	<i>Собеседование, контрольная работа</i>

		прикладных и теоретических задач теории строковых алгоритмов.	
	ПК-8.2. Иметь навыки разработки новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности	<p><u>Знать</u> как применять преобразования Барроуза-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; доказательство эквивалентности языков, задаваемых регулярными выражениями, недетерминированными и детерминированными конечными автоматами; алгоритмы минимизации детерминированных конечных автоматов: алгоритм Мура, алгоритм Хопкрофта; определение контекстно-свободной грамматики; сведение к нормальной форме Хомского; алгоритм СЮК;</p> <p><u>Уметь</u> использовать современные математические методы для анализа и разработки строковых алгоритмов.</p> <p><u>Владеть</u> навыками работы с современными библиотеками и наборами прикладных программ для строковых задач.</p>	<i>Контрольная работа</i>
	ПК-8.3. Иметь навыки управления разработкой и развитием новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности	<p><u>Знать</u> 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); основ-ые алгоритмы решения задачи нечеткого сопоставления: алгоритм 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;</p> <p><u>Владеть</u> способностью разрабатывать и</p>	<i>Собеседование</i>

		анализировать концептуальные и теоретические модели решаемых проблем и задач алгоритмов на строках; владение опытом реализации программных систем для решения практических задач теории строковых алгоритмов.	
--	--	---	--

### 3. Структура и содержание дисциплины

#### 3.1. Трудоемкость дисциплины

<b>Общая трудоемкость</b>	<b>3 ЗЕТ</b>
<b>Часов по учебному плану</b>	<b>108</b>
в том числе:	
аудиторные занятия (контактная работа):	33
- занятия лекционного типа	16
- занятия семинарского типа	
занятия лабораторного типа	16
- текущий контроль (КСР)	1
самостоятельная работа	75
<b>Промежуточная аттестация</b>	<b>зачет</b>

#### 3.2 Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины,  форма промежуточной аттестации по дисциплине	Всего (часы)	в том числе				
		контактная работа (работа во взаимодействии с преподавателем), часы, из них				Самостоятельная работа студента, часы
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Всего контактных часов	
Период строки. Префикс функция. Z-функция. Trie(бор) структура. Алгоритм Ахо-Корасик. Функция ошибок.	8	2		2	4	8
Суффиксный бор. Алгоритм Укконена. Задача о наибольшей общей подстроке. RMQ, SMQ и LCA структуры. Статический и динамический варианты.	9	2		2	4	9
Суффиксный массив. Построение суффиксного массива за линейное время. Построение LCP массива за линейное время. Применение техник хеширования для строковых задач. Алгоритм Рабина-Карпа.	9	2		2	4	9

Приближенное сопоставление строк с джокерами. Разные алгоритмы. Приближенное сопоставление строк по метрикам Хемминга и Левенштейна.	9	2		2	4	9
Задача о наибольшей общей подпоследовательности. Алгоритм Ханта-Сцимански, метод четырех русских. Построение сжатых индексов. Преобразование Барроуза-Уилера. Wavelet tree.	9	2		2	4	10
Конечные автоматы, введение. Минимальный автомат. Алгоритм Мура, Алгоритм Хопкрофта. Недетерминированные конечные автоматы. Регулярные выражения и регулярные языки.	9	2		2	4	10
Сведение регулярного выражения к НКА. Сведение ДКА к регулярному выражению. Лемма о накачке.	9	2		2	4	10
Контекстно свободные грамматики. Сведение к нормальной форме Хомского. Алгоритм СЮК.	9	2		2	4	10
Текущий контроль (КСР)	1				1	
<b>Итого</b>	<b>108</b>	<b>16</b>		<b>16</b>	<b>33</b>	<b>75</b>

Практические занятия (лабораторные занятия) организуются, в том числе в форме практической подготовки, которая предусматривает участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.

Практическая подготовка предусматривает: построение бордера строки; построение Z-функции; жонглирование суффиксного бора; создание и использование суффиксного массива.

На проведение практических занятий (лабораторных работ) в форме практической подготовки отводится 16 часов.

Практическая подготовка направлена на формирование и развитие:

- практических навыков в соответствии с профилем ОП: Разработка, тестирование, оптимизация программного обеспечения (ПО). Разработка технической документации на продукцию в сфере ИТ.
- компетенций – ПК-8: Способен к разработке новых алгоритмических, методических и технологических решений в конкретной сфере профессиональной деятельности (ПК-8.3: Имеет практический опыт анализа и интерпретации сложных информационных систем.).

Текущий контроль успеваемости реализуется в рамках занятий лабораторного типа, групповых или индивидуальных консультаций. Промежуточная аттестация проходит в традиционных формах (зачет).

#### 4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся осуществляется в виде работы с рекомендованной обязательной и дополнительной литературой, подготовке к лекциям, подготовке к зачету/экзамену и выполнения лабораторных работ. Контрольные вопросы и задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 5.2.

### 5. Фонд оценочных средств для промежуточной аттестации по дисциплине, включающий:

#### 5.1. Описание шкал оценивания результатов обучения по дисциплине

Уровень сформированности компетенций (индикатора достижения компетенций)	Шкала оценивания сформированности 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 незначительных ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок.	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 незначительными 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 владения материалом. 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	Продемонстрированы базовые навыки  при решении стандартных задач с некоторыми 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 оценки при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	Превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»
	Отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
	Очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
	Хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «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», ни одна из компетенций не сформирована на уровне «плохо»
	Плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

## 5.2 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения

### 5.2.1 Контрольные вопросы

	Вопрос	Код формируемой компетенции
1.	Сформулируйте понятие периода и бордера строки. Приведите линейные алгоритмы построения.	ПК-8
2.	Сформулируйте понятия префикс функции и Z-функции. Приведите линейные алгоритмы построения.	ПК-8
3.	Опишите алгоритм построения префиксного бора по набору образцов.	ПК-8
4.	Сформулируйте понятие функции ошибок.	ПК-8
5.	Опишите алгоритм Ахо-Корасик.	ПК-8
6.	Дайте определение суффиксного бора.	ПК-8
7.	Каким образом можно осуществить сжатие суффиксного бора до линейной памяти?	ПК-8
8.	Приведите алгоритм Укконена.	ПК-8
9.	Дайте определение суффиксного массива.	ПК-8
10.	Каким образом и за какое время можно решать задачу о подстроке пользуясь суффиксным массивом?	ПК-8
11.	Расскажите как строить суффиксный массив за линейное время.	ПК-8

### 5.2.2. Типовые тестовые задания для оценки сформированности компетенции ПК-8

Контрольная №1, вариант 2.

Дано слово «ДИССОНАНС».

- Постройте префиксный бор для данного слова.
- Постройте суффиксное дерево для данного слова.
- Постройте LCP-массив для данного слова.

Контрольная №2, вариант 1.

Даны строки «АБАСАБАСАБАБА» и «САБАБАСАБАСАБА»

- Найдите максимальную по длине общую подстроку этих строк.
- Найдите максимальную по длине общую подпоследовательность этих строк.
- Найдите редакционное расстояние между данными строками.

Контрольная №2, вариант 2.

Даны строки «АААББАБАБАСАБСАБ» и «ААББСАБСАББАБА»

- Найдите максимальную по длине общую подстроку этих строк.
- Найдите максимальную по длине общую подпоследовательность этих строк.
- Найдите редакционное расстояние между данными строками.

Контрольная №3

Реализуйте алгоритм Укконена построения суффиксного дерева. Проведите тесты корректности и быстродействия программы.



#### Контрольная №4

Реализуйте алгоритм Ахо-Корасик для построения префиксного бора. Проведите тесты корректности и быстродействия программы.

#### Контрольная №5

Реализуйте алгоритм Каркайнена-Сандерса для построения суффиксного массива. Проведите тесты корректности и быстродействия программы.

### 6. Учебно-методическое и информационное обеспечение дисциплины

- 1) а) основная литература:
  - 2) 1. Ден Гасфилд. Строки, деревья и последовательности в алгоритмах. Невский диалект. БХВ-Петербург. 2003. 656 стр. <http://sovremennik.ws/education/3007-stroki-derevja-i-posledovatelnosti-v-algoritmakh.html>
- б) дополнительная литература:
  - 3) 1. Crochemore, Maxime; Rytter, Wojciech (2003). Jewels of stringology. Text algorithms. River Edge, NJ: World Scientific. pp. 20–25. ISBN 981-02-4897-0. Zbl 1078.68151. <http://www.worldscientific.com/worldscibooks/10.1142/4838#t=oc>
- в) ссылки на интернет-ресурсы
  - 4) 1. <http://logic.pdmi.ras.ru/~yura/strings.html>
  - 5) 2. <https://www.coursera.org/learn/algorithms-on-strings>

### 7. Материально-техническое обеспечение дисциплины (модуля)

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой, оснащенные оборудованием и техническими средствами обучения: компьютерный класс, проектор, экран.

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

Учебная и научная литература, учебно-методические материалы, представленные в библиотечном фонде, в электронных библиотеках и на кафедре математического обеспечения и суперкомпьютерных технологий.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению 02.04.02 «Фундаментальная информатика и информационные технологии».

Автор к.ф.-м.н., Д. В. Грибанов

Заведующий кафедрой АГиДМ Н. Ю. Золотых

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики от 30 ноября 2022 года, протокол № 3.