

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ

Заведующий кафедрой
ПиИТ



проф. Махортов С.Д.,
10.03.2021

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.03. Алгоритмы на строках

- 1. Код и наименование направления подготовки/специальности:**
09.03.04 Программная инженерия
- 2. Профиль подготовки/специализация:**
Информационные системы и сетевые технологии
- 3. Квалификация выпускника:** бакалавр
- 4. Форма обучения:** очная
- 5. Кафедра, отвечающая за реализацию дисциплины:**
Программирования и информационных технологий (ПиИТ)
- 6. Составители программы:** Махортов Сергей Дмитриевич, д.ф.-м.н., доцент
- 7. Рекомендована** НМС ФКН, протокол № 5 от 10.03.2021.

отметки о продлении вносятся вручную)

8. Учебный год: 2024 / 2025

Семестр(ы): 8

9. Цели и задачи учебной дисциплины:

- изложить основы современных методов и алгоритмов эффективного вычисления образцов (паттернов) в строковых последовательностях и основы методологии анализа этих алгоритмов;
- рассмотреть типовые задачи обработки больших строковых последовательностей;
- научить студентов профессионально проектировать алгоритмы и структуры данных; вычислять и использовать оценки сложности алгоритмов с целью оптимизации, с учетом требований предметной области и потребностей пользователей;
- выработать способности и мотивацию к решению новых задач в рассматриваемой области, а также практические навыки применения полученных знаний.

10. Место учебной дисциплины в структуре ООП: Часть, формируемая участниками образовательных отношений (вариативная) блока Б1. Требуется предварительное знание основ математического анализа, дискретной математики, математической логики и теория алгоритмов, программирования. Предшествует дисциплинам: производственная практика, научно-исследовательская работа.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников):

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ПК-2	Способен выполнять тестирование ПО и анализировать полученные результаты	ПК-2.1	Разрабатывает план тестирования, выполняет построение тестовых случаев.	Знать: общие концепции алгоритмов на строках. Уметь: формулировать задачи оценки сложности. Владеть: математическим аппаратом для оценивания сложности алгоритмов.
ПК-4	Способен выполнять проектирование ПО	ПК-4.1 ПК-4.2	Разрабатывает и согласовывает архитектуру ПО с системным аналитиком. Проектирует структуры данных и баз данных.	Знать: критерии оценки алгоритмов, методы получения асимптотических оценок. Уметь: формулировать и доказывать оценки сложности. Владеть: подходами к исследованию алгоритмов, математическим аппаратом для оценивания сложности алгоритмов.
ПК-14	Способен организационно и технологически обеспечивать кодирование на языках программирования.	ПК-14.1	Обеспечивает соответствие кода и процесса кодирования принятым стандартам и технологиям.	Знать: конкретные алгоритмы на строках. Уметь: разрабатывать алгоритмы решения конкретных задач, сопоставлять различные алгоритмы. Владеть: подходами к исследованию алгоритмов на строках, соответствующим программным обеспечением.
ПК-21	Способен обрабатывать,	ПК-21.1	Обрабатывает полученные	Знать: конкретные алгоритмы на строках.

анализировать и оформлять результаты исследований и разработок под руководством специалиста более высокой квалификации	ПК-21.2	результаты исследований с использованием стандартных методов (методик). Применяет при обработке данных стандартное и оригинальное программное обеспечение.	Уметь: разрабатывать алгоритмы решения конкретных задач, сопоставлять различные алгоритмы. Владеть: подходами к исследованию алгоритмов на строках, соответствующим программным обеспечением.
--	---------	--	--

12. Объем дисциплины в зачетных единицах/час — 3 / 108.

Форма промежуточной аттестации зачет с оценкой.

13. Виды учебной работы

Вид учебной работы	Трудоемкость			
	Всего	По семестрам		
		№ 8	№ семестра	...
Аудиторные занятия	48	48		
в том числе:	лекции	-	-	
	практические	24	24	
	лабораторные	24	24	
Самостоятельная работа	60	60		
в том числе: курсовая работа (проект)	-	-		
Форма промежуточной аттестации (зачет с оценкой – 0 час.)	-	-		
Итого:	108	108		

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины
1. Лекции		
1.1	Введение. Структура строк	О строках, алгоритмах и оценках сложности. Асимптотические обозначения и сравнение функций. Грани строк. Z-блоки. Связь двух схем.
1.2	Задача о поиске образца	Алгоритм Кнута–Морриса–Пратта. Алгоритм Бойера–Мура. Алгоритм Карпа–Рабина. Алгоритм Shift-And.
1.3	Суффиксное дерево	Определение суффиксного дерева. Наивный алгоритм построения. Алгоритм с квадратичной сложностью. Алгоритм Укконена. Приложения суффиксных деревьев.
1.4	Суффиксный массив	Основные определения. Построение суффиксного массива. Приложения суффиксных массивов.
1.5	Оптимальное выравнивание строк	Расстояния между строками. Глобальное и локальное выравнивание. Алгоритм Вагнера-Фишера. Алгоритм Майерса.
2. Практические занятия		
3. Лабораторные работы		
3.1	Введение. Структура строк	О строках, алгоритмах и оценках сложности. Асимптотические обозначения и сравнение функций. Грани строк. Z-блоки. Связь двух схем.
3.2	Задача о поиске образца	Алгоритм Кнута–Морриса–Пратта. Алгоритм Бойера–Мура. Алгоритм Карпа–Рабина. Алгоритм Shift-And.
3.3	Суффиксное дерево	Определение суффиксного дерева. Наивный алгоритм построения. Алгоритм с квадратичной сложностью. Алгоритм Укконена. Приложения суффиксных деревьев.
3.4	Суффиксный массив	Основные определения. Построение суффиксного массива. Приложения суффиксных массивов.

	массив	
3.5	Оптимальное выравнивание строк	Расстояния между строками. Глобальное и локальное выравнивание. Алгоритм Вагнера-Фишера. Алгоритм Майерса.

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Введение. Структура строк		2	2	6	10
2	Задача о поиске образца		6	6	14	26
3	Суффиксное дерево		6	6	14	26
4	Суффиксный массив		6	6	12	24
5	Оптимальное выравнивание строк		4	4	14	22
Итого:			24	24	60	108

14. Методические указания для обучающихся по освоению дисциплины

Работа с конспектами лекций и презентационным материалом; выполнение практических заданий и тестов; выполнение лабораторных заданий; подготовка к заданиям текущей аттестации.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Томас Х. Алгоритмы: построение и анализ. 3-е издание / Х. Томас, Т. Кормен, Ч. Лейзерсон, Р. Ривест, К.Штайн. – М.: Вильямс, 2013. – 1328с.
2	<i>Окулов С.М.</i> Алгоритмы обработки строк. – М.: БИНОМ, 2013. – 255 с.
3	Гасфилд Д. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология / Пер. с англ. – СПб: Невский диалект, 2007. – 654 с.

б) дополнительная литература:

№ п/п	Источник
4	<i>Смит Б.</i> Методы и алгоритмы вычислений на строках / Пер. с англ. – М: Вильямс, 2006. – 496 с.
5	<i>Crochemore, M. and Wojciech, R.,</i> Jewels of Stringology, World Scientific, 2002, 310 p.

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
6	<i>Вяххи Н.И.</i> Алгоритмы в биоинформатике. Курс лекций [Электронный ресурс] / Computer Science клуб. – СПб, 2013. URL: http://compsciclub.ru/courses/algorithmsbioinformatics .
7	www.lib.vsu.ru – ЗНБ ВГУ
8	http://www.cs.vsu.ru/msd

16. Перечень учебно-методического обеспечения для самостоятельной работы (

№ п/п	Источник

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение)

Для реализации учебного процесса используются: бесплатные полнофункциональные инструментальные системы Qt Creator, Eclipse, PyCharm Community, Visual Studio Community; ресурс «Электронный университет» (<https://edu.vsu.ru/>).

18. Материально-техническое обеспечение дисциплины

1. Компьютерный класс №5 (ауд. 295). ПК-Intel-Core2 14 шт., рабочее место преподавателя: проектор, видеокоммутатор, специализированная мебель: доска маркерная 1 шт., столы 16 шт., стулья 33 шт. В классе находится точка доступа беспроводной сети для доступа в Интернет и к учебно-методическим материалам, расположенным на внутренних серверах факультета.

2. Компьютерный класс №7 (ауд. 316п). ПК на базе IntelCore2Duo 2,8ГГц, ОЗУ 2ГБ, диск 160Gb – 30 шт. Специализированная мебель: доска маркерная 1 шт., доска интерактивная 1 шт., столы 32 шт., стулья 64 шт.; рабочее место преподавателя: проектор, видеокоммутатор. В классе находится точка доступа беспроводной сети для доступа в Интернет и к учебно-методическим материалам, расположенным на внутренних серверах факультета.

19. Фонд оценочных средств

19.1. Перечень компетенций с указанием этапов формирования и планируемых результатов обучения

Код и содержание компетенции (или ее части)	Планируемые результаты обучения (показатели достижения заданного уровня освоения компетенции посредством формирования знаний, умений, навыков)	Этапы формирования компетенции (разделы (темы) дисциплины или модуля и их наименование)	ФОС* (средства оценивания)
ПК-2.1. Разрабатывает план тестирования, выполняет построение тестовых случаев.	Знать: общие концепции алгоритмов на строках.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Уметь: формулировать задачи оценки сложности.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Владеть: математическим аппаратом для оценивания сложности алгоритмов.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
ПК-4.1. Разрабатывает и согласовывает архитектуру ПО с системным аналитиком.	Знать: критерии оценки алгоритмов, методы получения асимптотических оценок.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Уметь: формулировать и доказывать оценки сложности.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Владеть: подходами к исследованию алгоритмов, математическим	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы

	аппаратом для оценивания сложности алгоритмов.		
ПК-4.2. Проектирует структуры данных и баз данных.	Знать: критерии оценки алгоритмов, методы получения асимптотических оценок.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Уметь: формулировать и доказывать оценки сложности.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Владеть: подходами к исследованию алгоритмов, математическим аппаратом для оценивания сложности алгоритмов.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
ПК-14.1. Обеспечивает соответствие кода и процесса кодирования принятым стандартам технологиям.	Знать: конкретные алгоритмы на строках.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Уметь: разрабатывать алгоритмы решения конкретных задач, сопоставлять различные алгоритмы.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Владеть: подходами к исследованию алгоритмов на строках, соответствующим программным обеспечением.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
ПК-21.1. Обрабатывает полученные результаты исследований использованием стандартных методов (методик).	Знать: конкретные алгоритмы на строках.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Уметь: разрабатывать алгоритмы решения конкретных задач, сопоставлять различные алгоритмы.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Владеть: подходами к исследованию алгоритмов на строках, соответствующим программным обеспечением.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
ПК-21.2. Применяет при обработке данных стандартное и оригинальное программное обеспечение.	Знать: конкретные алгоритмы на строках.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Уметь: разрабатывать алгоритмы решения конкретных задач, сопоставлять различные алгоритмы.	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив 5. Оптимальное выравнивание строк	Лабораторные работы
	Владеть: подходами к исследованию алгоритмов на строках, соответствующим	1. Введение. Структура строк 2. Задача о поиске образца 3. Суффиксное дерево 4. Суффиксный массив	Лабораторные работы

	программным обеспечением.	5. Оптимальное выравнивание строк	
Промежуточная аттестация			КИМ

19.2. Описание критериев и шкалы оценивания компетенций (результатов обучения) при промежуточной аттестации

Владение теоретическими основами дисциплины, способность иллюстрировать ответ примерами, данными научных исследований, применять теоретические знания для решения практических задач.

Для оценивания результатов обучения на зачете применяется 4-балльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Используются следующие показатели и их соотношения:

- уверенное владение теоретическими основами дисциплины, способность применять теоретические знания для решения практических задач, результаты выполнения всех заданий лабораторного практикума – «отлично»;
- хорошее владение теоретическими основами дисциплины, способность применять теоретические знания для решения практических задач, результаты выполнения большинства заданий лабораторного практикума – «хорошо»;
- неполное владение теоретическими основами дисциплины, затруднения в применении теоретических знаний для решения практических задач, результаты выполнения не менее 30% заданий лабораторного практикума – «удовлетворительно»;
- слабое владение теоретическими основами дисциплины, неспособность применять теоретические знания для решения практических задач, результаты выполнения менее 30% заданий лабораторного практикума – «неудовлетворительно».

19.3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения дисциплины

Текущий контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств: лабораторные работы. Перечень заданий для лабораторных работ строго соответствует темам занятий. Решение каждого задания должно быть доведено до компьютерной реализации.

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: собеседование по зачетным билетам (КИМ). Перечень вопросов к зачету:

1. Введение
 - 1.1. О строках, алгоритмах и оценках сложности
 - 1.2. Асимптотические обозначения и сравнение функций
2. Структура строк
 - 2.1. Грани строк
 - 2.2. Модифицированные массивы граней
 - 2.3. Z-блоки
 - 2.4. Связь двух схем
3. Задача о поиске образца
 - 3.1. Алгоритм Кнута–Морриса–Пратта
 - 3.2. О поиске в реальном времени
 - 3.3. Алгоритм Бойера-Мура
 - 3.3.1. Правило плохого символа
 - 3.3.2. Правило хорошего суффикса
 - 3.3.3. Общий алгоритм
 - 3.4. Алгоритм Карпа-Рабина
 - 3.5. Алгоритм Shift-And

- 4. Суффиксное дерево
 - 4.1. Основное определение и примеры
 - 4.2. Поиск на суффиксном дереве
 - 4.3. Наивный алгоритм построения суффиксного дерева
 - 4.4. Online подход
 - 4.4.1. Наивный алгоритм
 - 4.4.2. Алгоритм с квадратичной сложностью
 - 4.4.3. Линейный алгоритм Укконена
 - 4.5. Приложения суффиксных деревьев
 - 4.5.1. Множественный поиск
 - 4.5.2. Наибольшая общая подстрока
 - 4.5.3. Обобщенное суффиксное дерево
 - 4.5.4. Общие подстроки более чем двух строк
 - 4.5.5. Задача о подстроке для базы образцов
- 5. Суффиксный массив
 - 5.1. Основные определения
 - 5.2. Ускорение поиска
 - 5.3. Построение суффиксного массива
 - 5.3.1. Сортировка подсчетом
 - 5.3.2. Поразрядная сортировка
 - 5.3.3. Сортировка суффиксов
 - 5.4. Приложения суффиксных массивов
- 6. Расстояние между строками и выравнивание
 - 6.1. Глобальное и локальное выравнивание
 - 6.2. Алгоритм Вагнера-Фишера
 - 6.3. Алгоритм Майерса
 - 6.4. Применение к исследованию ДНК

19.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Оценка знаний, умений и навыков, характеризующая этапы формирования компетенций в рамках изучения дисциплины, осуществляется в ходе текущей и промежуточной аттестаций.

Текущая аттестация соответствует Положению о текущей аттестации обучающихся по программам высшего образования Воронежского государственного университета. Текущая аттестация проводится в форме выполнения и оценивания лабораторных работ. Решение каждого задания должно быть доведено до компьютерной реализации. Критерии оценивания приведены выше.

Промежуточная аттестация проводится в соответствии с Положением о промежуточной аттестации обучающихся по программам высшего образования.

Контрольно-измерительные материалы промежуточной аттестации включают в себя теоретические вопросы, позволяющие оценить уровень полученных знаний. При оценивании используются количественные шкалы оценок. Критерии оценивания приведены выше.