

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

УТВЕРЖДАЮ

Заведующий кафедрой
ПиИТ



/Махортов С.Д./
05.03.2024

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.ДВ.01.06.05. Алгоритмы на строках

- 1. Код и наименование направления подготовки/специальности:**
09.03.02 Информационные системы и технологии
- 2. Профиль подготовки/специализация:** все профили
- 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 05.03.2024.

отметки о продлении вносятся вручную)

8. Учебный год: 2027 / 2028

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 теории 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-5	Способен выполнять работы по созданию (модификации) и сопровождению информационных систем	ПК-5.2 ПК-5.3	Знает устройство и функционирование современных ИС, протоколы, интерфейсы и форматы обмена данными. Обеспечивает разработку и тестирование ИС на базе типовой ИС в соответствии с требованиями.	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

12. Объем дисциплины в зачетных единицах/час.(в соответствии с учебным планом) — 3 / 108.

Форма промежуточной аттестации – зачет с оценкой, курсовая работа.

13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость			
		Всего	По семестрам		
			№ 8	№ семестра	...
Аудиторные занятия		48	48		
в том числе:	лекции	24	24		
	практические	-	-		

	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.
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)				
		Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
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

(рекомендации обучающимся по освоению дисциплины: работа с конспектами лекций, презентационным материалом, выполнение практических заданий, тестов, заданий текущей аттестации и т.д.)

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 (список литературы оформляется в соответствии с требованиями ГОСТ и используется общая сквозная нумерация для всех видов источников)

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	Окулов С.М. 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	Смит Б. Методы и алгоритмы вычислений на строках / Пер. с англ. – М: Вильямс, 2006. – 496 с.
5	Crochemore, M. and Wojciech, R., Jewels of Stringology, World Scientific, 2002, 310 p.

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

№ п/п	Ресурс
6	Вяххи Н.И. 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

* Вначале указываются ЭБС, с которыми имеются договора у ВГУ, затем открытые электронно-образовательные ресурсы

16. Перечень учебно-методического обеспечения для самостоятельной работы (учебно-методические рекомендации, пособия, задачки, методические указания по выполнению практических (контрольных) работ и др.)

№ 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. Оценочные средства для проведения текущего контроля и промежуточной аттестаций

19.1. Текущий контроль

Текущая аттестация проводится в соответствии с Положением о текущей аттестации обучающихся по программам высшего образования Воронежского государственного университета. Текущая аттестация проводится в форме письменных работ (контрольные, выполнение практико-ориентированных заданий). Критерии оценивания приведены ниже.

19.1.1 Задания для текущего контроля успеваемости

Приведённые ниже задания рекомендуется использовать при проведении диагностических работ для оценки остаточных знаний по дисциплине

Тестовые задания – 1 балл

1. Что подразумевается в теории алгоритмов под временем работы алгоритма?
 - 1) Среднее время выполнения при множественных запусках.
 - 2) Число элементарных шагов, которые он выполняет.
 - 3) Время выполнения в миллисекундах на заранее оговоренном процессоре.
2. Согласно теории алгоритмов, от чего зависит время работы алгоритма?
 - 1) От компьютера, на котором он выполняется.
 - 2) От объема входных данных и их значений.
 - 3) Только от объема входных данных.
3. Какая из оценок времени работы алгоритма более информативна для практики?
 - 1) В лучшем случае.
 - 2) Средняя.
 - 3) В худшем случае.
 - 4) Выделить невозможно.
4. Какая из функций обозначает асимптотически точную оценку вычислительной сложности алгоритма?
 - 1) Ω .
 - 2) O .
 - 3) Θ .
5. Какая вычислительная сложность считается приемлемой для алгоритмов обработки больших строк?
 - 1) Экспоненциальная.
 - 2) Полиномиальная.
 - 3) Линейная.
6. Какова длина пустой строки?
 - 1) 0.
 - 2) Она не существует.
7. Какова вычислительная сложность «наивного» алгоритма поиска всех вхождений образца в текст?
 - 1) Линейная.
 - 2) Квадратичная.
 - 3) Кубическая.
8. Сколько непустых граней имеет строка АВААВАВААВААВ?
 - 1) 1.
 - 2) 2.
 - 3) 3.
9. Какая из перечисленных ниже схем описания структуры строк имеет более высокие выразительные возможности?
 - 1) Массив граней.
 - 2) Модифицированный массив граней.
 - 3) Массив Z-блоков.

- 4) Возможности равнозначны.
10. Что подвергается препроцессингу в алгоритме Кнута-Морриса-Пратта?
- 1) Искомый образец.
 - 2) Текст, в котором ищется образец.
11. Какой из перечисленных ниже алгоритмов имеет сублинейное среднее время работы?
- 1) Кнута-Морриса-Пратта.
 - 2) Бойера-Мура.
 - 3) Карпа-Рабина.
 - 4) Shift-And.
12. Какой из перечисленных ниже алгоритмов имеет модификацию online?
- 1) Кнута-Морриса-Пратта.
 - 2) Бойера-Мура.
 - 3) Карпа-Рабина.
 - 4) Shift-And.
13. Какие из упомянутых ниже алгоритмов можно отнести к категории получисленных?
- 1) Кнута-Морриса-Пратта.
 - 2) Бойера-Мура.
 - 3) Карпа-Рабина.
 - 4) Shift-And.
14. Какой из упомянутых ниже алгоритмов имеет вероятностное обоснование эффективности?
- 1) Кнута-Морриса-Пратта.
 - 2) Бойера-Мура.
 - 3) Карпа-Рабина.
 - 4) Shift-And.
15. Для быстрого поиска вхождений образца в текст какая из перечисленных ниже структур строится по самому тексту?
- 1) Массив граней.
 - 2) Массив Z-блоков.
 - 3) Суффиксное дерево.
16. Какова вычислительная сложность «наивного» алгоритма построения суффиксного дерева?
- 1) Линейная.
 - 2) Квадратичная.
 - 3) Кубическая.
17. Какова вычислительная сложность «наивного» последовательного алгоритма («с продлениями листьев») построения суффиксного дерева?
- 1) Линейная.
 - 2) Квадратичная.
 - 3) Кубическая.
18. Какова вычислительная сложность алгоритма Укконена построения суффиксного дерева?
- 1) Линейная.
 - 2) Квадратичная.
 - 3) Кубическая.
19. Какова вычислительная сложность «наивного» алгоритма построения суффиксного массива?
- 1) Линейная.
 - 2) $O(n \lg n)$.
 - 3) Квадратичная.
 - 4) $O(n^2 \lg n)$.
20. Какова вычислительная сложность алгоритма построения суффиксного массива на основе сортировки циклических сдвигов?
- 1) Линейная.

- 2) $O(n \lg n)$.
- 3) Квадратичная.
- 4) $O(n^2 \lg n)$.
21. Какова вычислительная сложность алгоритма построения суффиксного массива на основе суффиксного дерева?
- 1) Линейная.
- 2) $O(n \lg n)$.
- 3) Квадратичная.
- 4) $O(n^2 \lg n)$.

Ответы

Номер теста	Номер ответа
1	2
2	2
3	3
4	1
5	3
6	1
7	2
8	2
9	4
10	1
11	2
12	1
13	3, 4
14	3
15	3
16	2
17	3
18	1
19	4
20	2
21	1

Задания с кратким ответом – 2 балла

- Какая вычислительная сложность считается приемлемой для алгоритмов обработки больших строк?
- Какова длина пустой строки?
- Какова вычислительная сложность «наивного» алгоритма поиска всех вхождений образца в текст?
- Сколько непустых граней имеет строка АВААВАВААВААВ?
- Какова вычислительная сложность «наивного» алгоритма вычисления массива граней?
- Какова вычислительная сложность быстрого алгоритма вычисления массива граней?
- Какова вычислительная сложность «наивного» алгоритма вычисления массива Z-значений?
- Какова вычислительная сложность быстрого алгоритма вычисления массива Z-значений?
- Какова вычислительная сложность алгоритма Кнута-Морриса-Пратта?
- Что подвергается препроцессингу в алгоритме Кнута-Морриса-Пратта?
- Какой из известных Вам алгоритмов поиска образца в тексте имеет сублинейное среднее время работы?
- Какой из известных Вам алгоритмов поиска образца в тексте имеет модификацию online?
- Какие из известных Вам алгоритмов поиска образца в тексте можно отнести к категории получисленных?
- Какой из известных Вам алгоритмов поиска образца в тексте имеет вероятностное обоснование эффективности?

15. Какова вычислительная сложность «наивного» алгоритма построения суффиксного дерева?
16. Какова вычислительная сложность «наивного» последовательного алгоритма («с продлениями листьев») построения суффиксного дерева?
17. Какова вычислительная сложность алгоритма Укконена построения суффиксного дерева?
18. Какова вычислительная сложность алгоритма построения суффиксного массива на основе суффиксного дерева?
19. Какова вычислительная сложность алгоритма сортировки подсчетом?
20. Какова вычислительная сложность алгоритма поразрядной сортировки?

Ответы

Номер вопроса	Ответ
1	Линейная
2	0
3	Квадратичная
4	2
5	Кубическая
6	Линейная
7	Квадратичная
8	Линейная
9	Линейная
10	Образец
11	Бойера-Мура
12	Кнута-Морриса-Пратта
13	Карпа-Рабина, Shift-And
14	Карпа-Рабина
15	Квадратичная
16	Кубическая
17	Линейная
18	Линейная
19	Линейная
20	Линейная

Задания с развернутым ответом – 3 балла

1. Алгоритм вычисления массива граней. Описание и оценки сложности.
2. Алгоритм вычисления модифицированного массива граней. Описание и оценки сложности.
3. Алгоритм вычисления массива Z-блоков. Описание и оценки сложности.
4. Алгоритм Кнута-Морриса-Пратта. Описание и оценки сложности.
5. Алгоритм Бойера-Мура с правилом «плохого символа». Описание и оценки сложности.
6. Алгоритм Бойера-Мура с правилом «хорошего суффикса». Описание и оценки сложности.
7. Алгоритм Shift-And. Описание и оценки сложности.
8. Определение суффиксного дерева. Примеры.
9. «Наивный» алгоритм построения суффиксного дерева. Описание и оценки сложности.
10. Приложения суффиксных деревьев.
11. Обобщенное суффиксное дерево и его приложения.
12. Суффиксный массив. Основные определения и примеры.
13. Алгоритм сортировки «подсчетом». Описание и оценки сложности.
14. Алгоритм поразрядной сортировки. Описание и оценки сложности.
15. Алгоритм построения суффиксного массива. Описание и оценки сложности.

Критерии оценивания	Шкала оценок (в баллах)
Обучающийся отвечает на все вопросы правильно.	3
Обучающийся отвечает на все вопросы. Допускаются незначительные неточности.	2

Обучающийся отвечает не на все вопросы. Ответ не содержит грубых ошибок.	1
Обучающийся отвечает не на все вопросы. Присутствуют грубые ошибки.	0

19.2. Промежуточная аттестация

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 и практические задания, позволяющие оценить степень сформированности умений и навыков.

Для оценивания результатов обучения на зачете с оценкой используются следующие показатели: владение теоретическими основами дисциплины, способность иллюстрировать ответ примерами, 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».

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

Примерный перечень тем курсовых работ:

1. Задача о поиске образца
2. Суффиксное дерево
3. Суффиксный массив
4. Оптимальное выравнивание строк