

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

УТВЕРЖДАЮ

Заведующий кафедрой
ПиИТ
/Махортов С.Д./
10.03.2021

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.09. Алгоритмы на строках

- 1. Код и наименование направления подготовки/специальности:**
09.03.02 Информационные системы и технологии
- 2. Профиль подготовки/специализация:**
Программная инженерия в информационных системах
- 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 интеграцию программных модулей и компонент, выполнять верификацию программных продуктов | ПК-2.4 | Проверяет работоспособность программных продуктов | Знать: общие концепции алгоритмов на строках. Уметь: формулировать задачи оценки сложности. Владеть: математическим аппаратом для оценивания сложности алгоритмов |
| ПК-3 | Способен выполнять работы по созданию (модификации) и сопровождению информационных систем | ПК-3.4 | Разрабатывает код компонентов ИС и баз данных ИС | 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-4 | Способен проводить анализ требований к программному обеспечению, выполнять работы по проектированию программного обеспечения | ПК-4.4 | Описывает технологии обработки данных для возможности их использования в программной среде, включая вопросы параллельной обработки | 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-4.5 | Описывает применяемые математические | Знать: конкретные алгоритмы на строках. Уметь: разрабатывать алгоритмы решения |

| | | | | |
|--|--|--|---|--|
| | | | методы и алгоритмы функционирования для компонентов программных средств | 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. Трудоемкость по видам учебной работы

| Вид учебной работы | Трудоемкость | | | |
|---|--------------|--------------|------------|-----|
| | Всего | По семестрам | | |
| | | № 8 | № семестра | ... |
| Аудиторные занятия | 72 | 72 | | |
| в том числе: | лекции | 36 | 36 | |
| | практические | - | - | |
| | лабораторные | 36 | 36 | |
| Самостоятельная работа | 36 | 36 | | |
| в том числе: курсовая работа (проект) | - | - | | |
| Форма промежуточной аттестации (зачет – 0 час.) | - | - | | |
| Итого: | 108 | 108 | | |

13.1. Содержание дисциплины

| п/п | Наименование раздела дисциплины | Содержание раздела дисциплины | Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК |
|--------------------------------|---------------------------------|---|--|
| 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. | Онлайн-курс «Алгоритмы на строках» |
| 1.2 | Задача о поиске образца | Алгоритм Кнута–Морриса–Пратта. Алгоритм Бойера–Мура. Алгоритм Карпа–Рабина. Алгоритм Shift-And. | Онлайн-курс «Алгоритмы на строках» |
| 1.3 | Суффиксное дерево | Определение суффиксного дерева. Наивный алгоритм построения. Алгоритм с квадратичной сложностью. Алгоритм Укконена. Приложения суффиксных деревьев. | Онлайн-курс «Алгоритмы на строках» |
| 1.4 | Суффиксный массив | Основные определения. Построение суффиксного массива. Приложения суффиксных массивов. | Онлайн-курс «Алгоритмы на строках» |
| 1.5 | Оптимальное выравнивание строк | Расстояния между строками. Глобальное и локальное выравнивание. Алгоритм Вагнера–Фишера. Алгоритм Майерса. | Онлайн-курс «Алгоритмы на строках» |
| 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) | | | | Всего |
|-------|--|----------------------|--------------|--------------|------------------------|-------|
| | | Лекции | Практические | Лабораторные | Самостоятельная работа | |
| 1 | Введение. Структура строк | 4 | | 4 | 4 | 12 |
| 2 | Задача о поиске образца | 10 | | 10 | 10 | 30 |
| 3 | Суффиксное дерево | 8 | | 8 | 8 | 24 |
| 4 | Суффиксный массив | 8 | | 8 | 8 | 24 |
| 5 | Оптимальное выравнивание строк | 6 | | 6 | 6 | 18 |
| | Итого: | 36 | | 36 | 36 | 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., <i>Jewels of Stringology</i> , 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://compsclub.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 используется бесплатная полнофункциональная инструментальная система Qt Creator 4.5.0, 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 текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

| № п/п | Наименование раздела дисциплины (модуля) | Компетенция(и) | Индикатор(ы) достижения компетенции | Оценочные средства |
|--|--|----------------------|-------------------------------------|--|
| 1 | Введение. Структура строк | ПКВ-2, ПКВ-4 | ПКВ-2.1, ПКВ-4.2 | Практические задания и тесты на порталах http://www.cs.vsu.ru/msd , https://edu.vsu.ru |
| 2 | Задача о поиске образца | ПКВ-2, ПКВ-4, ПКВ-14 | ПКВ-2.1, ПКВ-4.2, ПКВ-14.1 | Практические задания и тесты на порталах http://www.cs.vsu.ru/msd , https://edu.vsu.ru |
| 3 | Суффиксное дерево | ПКВ-2, ПКВ-4, ПКВ-14 | ПКВ-2.1, ПКВ-4.2, ПКВ-14.1 | Практические задания и тесты на порталах http://www.cs.vsu.ru/msd , https://edu.vsu.ru |
| 4 | Суффиксный массив | ПКВ-2, ПКВ-4, ПКВ-14 | ПКВ-2.1, ПКВ-4.2, ПКВ-14.1 | Практические задания и тесты на порталах http://www.cs.vsu.ru/msd , https://edu.vsu.ru |
| 5 | Оптимальное выравнивание строк | ПКВ-2, ПКВ-4, ПКВ-14 | ПКВ-2.1, ПКВ-4.2, ПКВ-14.1 | Практические задания и тесты на порталах http://www.cs.vsu.ru/msd , https://edu.vsu.ru |
| Промежуточная аттестация форма контроля – зачет с оценкой | | | | Практические задания и тесты на порталах http://www.cs.vsu.ru/msd , https://edu.vsu.ru |

20. Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1. Текущий контроль успеваемости

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

20.2. Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: собеседование по билетам к зачету.

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

Для оценивания результатов обучения на зачете 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».