

Белорусский государственный университет



Проректор по учебной работе и
образовательным инновациям
О.И. Чуприс
2019 г.

Регистрационный № УД- 6950/уч.

СПЕЦИАЛЬНЫЕ СТРУКТУРЫ ДАННЫХ

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

1-31 80 09 Прикладная математика и информатика

профилизации:

Алгоритмы и системы обработки больших данных

Аналитическая логистика

Компьютерный анализ данных

Математическая кибернетика

2019 г.

Учебная программа составлена на основе ОСВО 1-31 80 09-2019 и учебных планов G31-072/уч., G31-073/уч., G31-074/уч., G31-075/уч. от 11.04.2019 г.

СОСТАВИТЕЛЬ:

А.А. Толстик – старший преподаватель кафедры вычислительной математики факультета прикладной математики и информатики Белорусского государственного университета.

РЕЦЕНЗЕНТЫ:

О.И. Мельников – профессор кафедры математической кибернетики механико-математического факультета БГУ, доктор педагогических, доцент;

С.В. Жгировский – инженер-программист ООО «ЯндексБел».

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой дискретной математики и алгоритмики факультета прикладной математики и информатики БГУ

(протокол № 15 от 18 апреля 2019 года);

Научно-методическим Советом БГУ

(протокол № 5 от 28 июня 2019 года).

Заведующий кафедрой
дискретной математики и алгоритмики



В.М. Котов



ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Учебная дисциплина «Специальные структуры данных» знакомит студентов магистратуры с теоретическими аспектами и принципами проектирования и реализации современных эффективных структур данных, примерах использования в программных комплексах и системах.

Цель учебной дисциплины – создание базы для использования современных библиотек в различных языках программирования, умения адаптировать существующие алгоритмы и структуры данных для конкретных прикладных задач. Развитие умения оценивать трудоемкость операций с структурами данных, выполняться оценку в худшем случае и амортизированную оценку операций, использовать вероятностную оценку трудоемкости операций.

При изложении материала учебной дисциплины важно показать многообразие современных структур данных, классифицируя их по таким критериям, как время работы операций, объем используемой памяти, возможность использовать в многопоточном режиме.

Задачи учебной дисциплины:

1. Изучение специальных структур данных: дерево интервалов, битовые множества, левосторонние кучи, персистентные версии различных структур данных.

2. Изучение методов организации эффективного перебора решений задачи, оптимизация метода, используя различные структуры данных.

3. Использование изученных специальных структур данных в качестве блоков при разработке различных программных продуктов и сервисов.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием (магистра).

Учебная дисциплина относится к государственному компоненту и входит в модуль «Алгоритмические аспекты компьютерных наук».

Программа составлена с учетом **межпредметных связей** с учебными дисциплинами. Основой для изучения учебной дисциплины являются следующие учебные дисциплины первой ступени высшего образования: «Теория алгоритмов», «Программирование», «Теория вероятностей и математическая статистика», «Теория графов» и «Дискретная математика и математическая логика». Знания, полученные в учебной дисциплине, используются при изучении дисциплин вузовского компонента «Вероятностные алгоритмы и структуры данных», «Алгоритмы обработки текстов», «Алгоритмы во внешней памяти» модуля «Специализированные алгоритмы и структуры данных».

Требования к компетенциям

Освоение учебной дисциплины «Специальные структуры данных» должно обеспечить формирование следующих универсальных, углубленных профессиональных компетенций.

универсальные компетенции:

УК-4. Обладать способностью проектирования и использования абстрактных моделей и структур.

углубленные профессиональные компетенции:

УПК-4. Оценивать эффективность алгоритмов решения прикладных задач.

В результате изучения дисциплины студент магистратуры должен:

знать:

- понятие трудоемкости алгоритма;
- основные подходы при разработке эффективных алгоритмов;
- способы организации структур данных и технологию их использования;
- виды поисковых деревьев;
- методы построения эффективных переборных алгоритмов.

уметь:

- адаптировать структуры данных при разработке алгоритма решения задачи;
- реализовывать различные структуры данных;
- декомпозировать задачу для более эффективного использования известных алгоритмов и структур данных;

владеть:

- навыками адаптации, реализации и использования структур данных;
- навыками использования стандартных библиотек алгоритмов и структур данных.

Структура учебной дисциплины

Дисциплина изучается в 1-ом семестре. Всего на изучение учебной дисциплины «Специальные структуры данных» отведено:

– для очной формы получения высшего образования – 198 часов, в том числе 60 аудиторных часов, из них: лекции – 20 часов, семинарские занятия – 20 часов, лабораторные занятия – 20 часов.

Трудоемкость учебной дисциплины составляет 6 зачетных единиц.

Форма текущей аттестации по учебной дисциплине – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Интервальные структуры данных

Тема 1.1. Введение. Оценка трудоемкости операций со структурами данных

Абстрактные структуры данных. Понятие интерфейса структуры данных. Формирование цели использования и основных показателей эффективности структур данных.

Оценка времени работы в худшем случае. Амортизированная оценка времени работы. Метод потенциалов. Оценка времени работы на случайных входных данных. Примеры структур данных с балансировкой времени работы между различными операциями. Примеры структур данных с балансировкой между объемом используемой памяти и временем работы.

Тема 1.2. Дерево отрезков

Структура данных «дерево отрезков». Особенности реализации в динамической памяти и на массиве ячеек. Декомпозиция отрезка запроса на элементарные отрезки. Операции изменения элемента. Операции запросов: значение элемента, сумма на интервале, минимум на интервале, количество различных элементов на интервале. Метод частичного каскадирования. Оценка трудоемкости операций.

Тема 1.3. Дерево отрезков. Интервальная модификация

Особенности применения операций интервальной модификации: установка значений на интервале, прибавление значений на интервале. Операции запросов: значение элемента, сумма на интервале, минимум на интервале, количество различных элементов на интервале. Оценка трудоемкости операций. Применение в решении геометрических задач.

Раздел 2. Декомпозиция данных

Тема 2.1. Битовое представление множества

Реализация битового представления множества и операций с ним. Оценка эффективности операций. Решение системы линейных уравнений. Поиск ранга матрицы. Поиск транзитивного замыкания в графе. Алгоритм поиска в ширину в ориентированном графе.

Тема 2.2. Декомпозиция запросов

Общий принцип техники декомпозиции запросов, «простые» операции, перестроение структуры данных. Оценка эффективности метода. Определение оптимального порогового значения. Решение задачи подсчета треугольников в графе. Решение задачи поиска наименьшего общего предка в динамически достраиваемом дереве.

Тема 2.3. Эффективная организация вычислений в переборных алгоритмах

Метод большого и малого шага. Решение задачи дискретного логарифмирования с трудоемкостью $O(n^{0.5})$. Поиск длины цикла. Метод встречи посередине. Решение задачи поиска подмножества с заданным критерием на сумму элементов. Подсчет количества гамильтоновых циклов. Подсчет количества пар предметов с фиксированным набором признаков.

Раздел 3. Персистентные структуры данных

Тема 3.1. Классификация. Общие принципы построения. Общие методы построения персистентных структур данных

Краткая характеристика персистентных структур данных. Классификация персистентности структур данных: частичная, полная, конфлюэнтная, функциональная. Метод полного копирования. Метод копирования пути. Метод «толстых» узлов.

Тема 3.2. Персистентный стек, список, очередь

Построение полностью персистентного стека на основе корневого дерева. Преобразование списка в частично персистентный за константное время. Реализация интерфейса очереди на двух стеках с сохранением амортизационной оценки. Реализация интерфейса очереди на шести стеках с сохранением гарантированной оценки. Реализация персистентной очереди на пяти стеках.

Тема 3.3 Персистентная очередь с приоритетом, дерево отрезков, бинарное поисковое дерево

Построение персистентного дерева отрезков методом копирования пути. Построение персистентного бинарного поискового дерева методом копирования пути. Левосторонние кучи. Биномиальные кучи. Построение персистентной очереди с приоритетами.

Тема 3.4 Построение частично персистентных и полностью персистентных структур данных

Недостатки и преимущества различных версий персистентных структур данных. Общий метод построения частично персистентных структур данных. Эйлеров обход дерева. Поддержание порядка в списке. Получение полностью персистентных структур данных.

Раздел 4. Дополнительные главы

Тема 4.1. Специализированные прикладные структуры данных

Списки с пропусками. LSM дерево. 2-3 дерево. R , R^+ , R^* дерево. Куча Бродала-Окасаки. Динамическая связность.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования

| № п/п | Название раздела, темы | Количество часов | | | | Количество часов УСР | Форма контроля знаний |
|-----------|---------------------------------------------------------------|------------------|------------------------------------------|-------------------------|------|------------------------------------------------------------------------------------------------------------|-----------------------|
| | | Аудиторные | | | | | |
| | | Лекции | Практические и семинарские занятия | Лабораторные занятия | Иное | | |
| 1. | Интервальные структуры данных | 6 | 2 | 6 | | | |
| 1.1 | Введение. Оценка трудоемкости операций со структурами данных. | 2 | | 2 | | Отчет лабораторным работам с их устной защитой. | |
| 1.2 | Дерево отрезков | 2 | | 2 | | Отчет лабораторным работам с их устной защитой. | |
| 1.3 | Дерево отрезков. Интервальная модификация | 2 | 2 | 2 | | Отчет лабораторным работам с их устной защитой. Контрольная работа. | |
| 2. | Декомпозиция данных | 6 | 4 | 6 | | | |
| 2.1 | Битовое представление множества | 2 | 4 | 2 | | Отчет лабораторным работам с их устной защитой. Выступление с докладом на семинаре. | |
| 2.2 | Декомпозиция запросов | 2 | | 2 | | Отчет лабораторным работам с их устной защитой. | |
| 2.3 | Эффективная организация вычислений в переборных алгоритмах | 2 | | 2 | | Отчет лабораторным работам с их устной защитой. Выступление с докладом на семинаре. Контрольная работа. | |

| | | | | | | | |
|-----------|-------------------------------------------------------------------------------------------------|----------|-----------|----------|--|--|-------------------------------------------------------------------------------------|
| 3. | Персистентные структуры данных | 8 | 2 | 8 | | | |
| 3.1 | Классификация. Общие принципы построения. Общие методы построения персистентных структур данных | 2 | | | | | Устный опрос |
| 3.2 | Персистентный стек, список, очередь | 2 | | | | | Коллоквиум |
| 3.3 | Персистентная очередь с приоритетом, дерево отрезков, бинарное поисковое дерево | 2 | | 4 | | | Отчет лабораторным работам с их устной защитой. Контрольная работа. |
| 3.4 | Построение частично персистентных и полностью персистентных структур данных | 2 | 2 | 4 | | | Отчет лабораторным работам с их устной защитой. Выступление с докладом на семинаре. |
| 4. | Дополнительные главы | | 12 | | | | |
| 4.1 | Специализированные прикладные структуры данных | | 12 | | | | Выступление с докладом на семинаре. |

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Теория алгоритмов: учеб. пособие / П.А. Иржавский, В.М. Котов, А.Ю. Лобанов, Ю.Л. Орлович, Е.П. Соболевская – Минск : БГУ, 2013. – 159 с. Допущено Министерством образования Республики Беларусь в качестве учебного пособия для магистрантов учреждения высшего образования по специальностям «Прикладная математика и информатика», «Теоретические основы информатики».
2. Сборник задач по теории алгоритмов: учеб.-метод. пособие / В.М. Котов, Ю.Л. Орлович, Е.П. Соболевская, С.А. Соболев – Минск : БГУ, 2017.- 183с.
3. В.М. Котов, Е. П. Соболевская, А. А. Толстиков. «Алгоритмы и структуры данных»: учеб. пособие (допущено Министерством образования Республики Беларусь в качестве учебного пособия для студентов учреждений высшего образования по специальностям «Прикладная математика», «Информатика», «Актуарная математика») Минск: БГУ, 2011г. – 267 с. – (Классическое университетское издание).
4. Кормен Т., Лейзерсон Ч., Ривест Р. Алгоритмы: построение и анализ, 2-е издание.: Пер. с англ. – М.: Издательский дом «Вильямс», 2005. – 1296с.
5. Mark de Berg, Marc van Kreveld, Mark Overmars, Otfried Schwarzkopf. More Geometric Data Structures. Computational Geometry: algorithms and applications (3rd ed.). Springer-Verlag Berlin Heidelberg New York, 2008. – 386 p.

Перечень дополнительной литературы

1. Ахо Альфред В., Хопкрофт Джон Э., Ульман Джеффри Д. Структуры данных и алгоритмы. М.: Издательский дом «Вильямс», 2000. – 384 с.
2. D. Shanks. Class number, a theory of factorization and genera. In Proc. Symp. Pure Math. 20, pages 415 – 440. AMS, Providence, R.I., 1971.
3. Haim Kaplan. Persistent data structures. Handbook on Data Structures and Applications. CRC Press, 2001. – 27 p.
4. William Pugh. Skip Lists: A Probabilistic Alternative to Balanced Trees / Workshop on Algorithms and Data Structures. Springer Berlin Heidelberg, 1989; Communications of the ACM CACM Homepage archive Volume 33 Issue 6, June 1990: 668-676.
5. O'Neil, Patrick E.; Cheng, Edward; Gawlick, Dieter; O'Neil, Elizabeth. The log-structured merge-tree (LSM-tree). Acta Informatica, 1996. 33 (4): 351–385.

6. Willard, Dan E. Log-logarithmic worst-case range queries are possible in space $\Theta(N)$. Information Processing Letters, 1993. Elsevier. 17 (2): 81–84.
7. Bose, Prosenjit; Douïeb, Karim; Dujmović, Vida; Howat, John; Morin, Pat, Fast Local Searches and Updates in Bounded Universes, Proceedings of the 22nd Canadian Conference on Computational Geometry (CCCG2010): 261–264.
8. Guttman, A. R-Trees: A Dynamic Index Structure for Spatial Searching. Proceedings of the 1984 ACM SIGMOD international conference on Management of data - SIGMOD '84. p. 47.
9. Beckmann, N.; Kriegel, H. P.; Schneider, R.; Seeger, B. The R*-tree: an efficient and robust access method for points and rectangles. Proceedings of the 1990 ACM SIGMOD international conference on Management of data - SIGMOD '90. p. 322.
10. Sleator, D. D.; Tarjan, R. E. A Data Structure for Dynamic Trees. Proceedings of the thirteenth annual ACM symposium on Theory of computing - STOC '81. p. 114.
11. Jacob Holm, Kristian de Lichtenberg, Mikkel Thorup: Poly-logarithmic deterministic fully dynamic algorithms for connectivity, minimum spanning tree, 2-edge, and biconnectivity. J. ACM 48(4): 723-760.
12. Brodal, Gerth S. Worst-Case Efficient Priority Queues, Proc. 7th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 52–58.

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

Для диагностики компетенции в рамках учебной дисциплины рекомендуется использовать следующие формы:

1. Устная форма: устный опрос, доклады на семинарских занятиях, коллоквиум.
2. Письменная форма: контрольные работы.
3. Устно-письменная форма: отчеты по лабораторным работам с их устной защитой, оценивание на основе проектного метода.

Формой текущей аттестации по дисциплине «Специальные структуры данных» учебным планом предусмотрен экзамен.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний студентов по дисциплине.

Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний в рейтинговую оценку:

- отчёты по лабораторным работам с их устной защитой – 50 %;
- контрольные работы – 10 %;
- коллоквиум – 10 %;
- устный опрос – 10%;
- выступление с докладом на семинаре – 20%.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и экзаменационной оценки с учетом их весовых коэффициентов. Вес оценка по текущей успеваемости составляет 30 %, экзаменационная оценка – 70%.

Примерная тематика лабораторных занятий

Занятие № 1. Реализация стека/очереди/дека с поддержанием минимума.

Занятие № 2. Реализация СД «Дерево отрезков» по заданному интерфейсу.

Занятие № 3. Реализация СД «Дерево отрезков» по заданному интерфейсу с интервальной модификацией.

Занятие № 4. Реализация поиска в ширину в ориентированном графе с использованием битовых представлений.

Занятие № 5. Реализация алгоритма подсчета треугольников в неориентированном графе.

Занятие № 6. Реализация алгоритма решения задачи о рюкзаке методом разделения предметов на два множества.

Занятие № 7. Реализация персистентной версии структуры данных стек.

Занятие № 8. Реализация персистентной версии структуры данных дерево отрезков.

Занятие № 9. Реализация алгоритма сериализации дерева в последовательность вершин (Эйлеров обход).

Занятие № 10. Реализация алгоритма десериализации дерева в последовательность вершин (Эйлеров обход).

Примерная тематика семинарских занятий

Занятие № 1. Двумерное дерево отрезков.

Занятие № 2. Быстрый цифровой бор (X-fast trie).

Занятие № 3. Сверхбыстрый цифровой бор (Y-fast trie).

Занятие № 4. Link/cut дерево.

Занятие № 5. Списки с пропусками.

Занятие № 6. LSM дерево.

Занятие № 7. 2-3 дерево.

Занятие № 8. R, R+, R* дерево.

Занятие № 9. Куча Бродала-Окасаки.

Занятие № 10. Динамическая связность.

Описание инновационных подходов и методов к преподаванию учебной дисциплины (эвристический, проективный, практико-ориентированный)

При организации образовательного процесса большинства практических занятий используется практико-ориентированный подход, который предполагает освоение содержания учебного материала через решение практических задач, а также приобретение навыков эффективного выполнения разных видов профессиональной деятельности.

Кроме этого, при организации образовательного процесса используется комбинация методов группового обучения, проектного обучения и учебной дискуссии. Комбинация методов предполагает: ориентацию на генерирование идей, приобретение навыков для решения исследовательских, творческих и коммуникационных задач, появление нового уровня понимания изучаемой темы, применение знаний (теорий, концепций) при решении проблем, определение способов их решения.

Методические рекомендации по организации самостоятельной работы обучающихся, подготовка к экзамену

Для организации самостоятельной работы студентов магистратуры по учебной дисциплине следует использовать современные информационные технологии: системы автоматического тестирования (образовательный портал InsightRunner – www.acm.bsu.by), разместить в сетевом доступе комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, презентации лекций, методические указания к практическим занятиям, электронные версии домашних заданий, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в том числе вопросы для подготовки к экзамену, задания, вопросы для самоконтроля, список рекомендуемой литературы, информационных ресурсов и др.).

Примерный перечень вопросов к экзамену

1. Оценка трудоемкости алгоритма. Трудоемкость в худшем случае.
2. Оценка трудоемкости алгоритма. Амортизированная оценка.
3. Оценка трудоемкости алгоритма. Метод потенциалов.
4. Понятие интерфейса абстрактной структуры данных. Примеры интерфейсов с балансировкой времени работы методов и объемом используемой памяти.

5. Оценка времени работы операций инкрементирования счетчика (изменение битов в счетчике).
6. Понятие структуры данных «Дерево отрезков». Особенности реализации в динамической и статической памяти.
7. Метод и оценка разбиения запроса на элементарные отрезки в «Дереве отрезков».
8. Метод «ленивого» проталкивания модификаций в «Дереве отрезков».
9. Операция изменения элемента и операция подсчета суммы на отрезке в задаче поиска суммы на отрезке.
10. Операция изменения элемента и операция нахождения минимального элемента в задаче поиска минимума на отрезке.
11. Операция множественного изменения значений (установка, прибавление) на отрезке одним запросом. Подсчет суммы на отрезке.
12. Поиск отрезка с максимальной суммой в последовательности.
13. Операция модификации элемента в задаче поиска отрезка с максимальной суммой.
14. Метод хранения большого количества информации в каждой вершине дерева. Подсчет количества элементов меньше заданного на участке массива.
15. Битовые представления множеств.
16. Битовые представления множеств. Подсчет количества гамильтоновых циклов.
17. Битовые представления множеств. Подсчет количества пар предметов с фиксированным набором признаков.
18. Битовые представления множеств. Транзитивное замыкание в ориентированном графе.
19. Битовые представления множеств. Поиск в ширину в ориентированном графе.
20. Решение системы линейных уравнений, используя битовые представления множеств.
21. Вычисление ранга матрицы, используя битовые представления множеств.
22. Метод декомпозиции данных и запросов. Подсчет треугольников в графе.
23. Метод декомпозиции данных и запросов. Подсчет высоты вершины в перестраиваемом дереве специального вида.
24. Метод большого и малого шага. Решение задачи дискретного логарифмирования.
25. Метод большого и малого шага. Определение длины цикла последовательности.
26. Персистентные структуры данных. Определения. Классификация. Примеры.
27. Персистентный стек.
28. Персистентный список.
29. Персистентная очередь.
30. Персистентная очередь с приоритетом.

31. Персистентное дерево отрезков. Метод копирования пути.
32. Персистентное бинарное дерево. Метод копирования пути.
33. Общий метод построения частично персистентных структур данных.
34. Эйлеров обход дерева.
35. Поддержание порядка в списке.
36. Получение полностью персистентных структур данных.

Рекомендуемая тематика контрольных работ

1. *Контрольная работа № 1.* Интервальные структуры данных.
2. *Контрольная работа № 2.* Декомпозиция данных.
3. *Контрольная работа № 3.* Персистентные структуры данных.
4. *Коллоквиум.* Способы организации специальных структур данных и технология их использования.

Текущий контроль знаний проводится в соответствии с учебно-методической картой дисциплины.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ

| Название учебной дисциплины, с которой требуется согласование | Название кафедры | Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине | Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола) |
|---------------------------------------------------------------|-------------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| Вероятностные алгоритмы и структуры данных | Дискретной математики и алгоритмики | Нет | Оставить содержание учебной дисциплины без изменения, протокол № 15 от 18 апреля 2019 г. |
| Алгоритмы обработки текстов | Дискретной математики и алгоритмики | Нет | Оставить содержание учебной дисциплины без изменения, протокол № 15 от 18 апреля 2019 г. |
| Алгоритмы во внешней памяти | Дискретной математики и алгоритмики | Нет | Оставить содержание учебной дисциплины без изменения, протокол № 15 от 18 апреля 2019 г. |

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ

на ____ / ____ учебный год

| №№ Пп | Дополнения и изменения | Основание |
|----------|------------------------|-----------|
| | | |

Учебная программа пересмотрена и одобрена на заседании кафедры дискретной математики и алгоритмики (протокол № ____ от _____ 20__ г.)

Заведующий кафедрой

(ученая степень, звание)

(подпись)

(И.О. Фамилия)

УТВЕРЖДАЮ

Декан факультета

(ученая степень, звание)

(подпись)

(И.О.Фамилия)