

Программа учебной дисциплины «Алгоритмы для иерархий памяти»

Утверждена

Академическим советом ООП

Протокол № 2.3 – 09/3008 - 01

от «30» августа 2018 г.

Автор	Бабенко М.А. Колесниченко И.И.
Число кредитов	5
Контактная работа (час.)	60
Самостоятельная работа (час.)	130
Курс	4
Формат изучения дисциплины	Без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Алгоритмы для иерархий памяти» являются ознакомление студентов с основными принципами построения алгоритмов для работы с данными, которые не помещаются в оперативную память компьютера.

В результате освоения дисциплины студент должен:

знать:

Основные алгоритмы для работы с данными, находящимися во внешней памяти, важнейшие cash-oblivious алгоритмы. Устройство SDD-дисков. Архитектуру компьютера, устройство жёсткого диска и процессорного кэша, особенности их устройства в системе Linux.

уметь:

Понять поставленную задачу; реализовать собственный алгоритм для работы с данными во внешней памяти; строить различные структуры данных во внешней памяти; работать с графами во внешней памяти; самостоятельно находить способы выполнения поставленных задач, в том числе и нестандартных, и проводить их анализ; самостоятельно видеть следствия полученных результатов. Делать оценки производительности алгоритмов.

владеть:

Навыками освоения большого объема информации и решения задач (в том числе, сложных); навыками самостоятельной работы и освоения новых дисциплин; культурой постановки, анализа и решения задач, возникающих при работе с большими данными.

Изучение дисциплины «Алгоритмы для иерархий памяти» базируется на следующих дисциплинах:

- алгоритмы и структуры данных.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- обладать навыками программирования на языке C++;
- знать классические алгоритмы и структуры данных.

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Модель вычислений во внешней памяти. Сортировка во внешней памяти.

Задача List Ranking. Рандимизированное и детерминированное решение. Задача о вершинной раскраске графа степени k в $k+1$ цвет.

(a,b)-деревья. Буфферизированные деревья.

Кучи во внешней памяти. Графы во внешней памяти. Обход в ширину. Поиск связанных компонент алгоритмом Борувки.

Ускоренный обход в ширину во внешней памяти.

Кеширование. OPT- и LRU-стратегии вытеснение, соотношение качества их работы.

Cache oblivious алгоритмы. Cache oblivious бинарный поиск. van Emde Boas layout.

Cache-oblivious транспонирование матриц. COLA. Fractional cascading.

Введение в streaming. Задача подсчета количества вхождений значения в потоке. Модели turnstile, strong turnstile, cash-register. Count-min sketch. Алгоритм Misra-Gries.

Подсчет количества различных элементов в потоке. Алгоритм AMS. Неравенство Чернова и медианный трюк. Алгоритм VJKST.

Поиск порядковых статистик за ограниченное число проходов. Алгоритм Munro-Paterson.

Однопроходный алгоритм поиска медианы.

Count Sketch. Линейные скетчи. Понижение размерности и лемма Джонсона-Линденштрауса. Tug-of-War Sketch.

Locality sensitive hashing. Min-wise семейства перестановок. LSH для коэффициента Жаккара. LSH для эвклидового угла между векторами.

III. ОЦЕНИВАНИЕ

Текущий контроль - домашняя работа в первом модуле, домашняя работа во втором модуле, контрольная работа в первом модуле, контрольная работа во втором модуле.

Итоговый контроль – письменный экзамен (120 мин.) Итоговая оценка вычисляется следующим образом: $0,2 * \text{оценка за домашнее задание No1} + 0,2 * \text{оценка за домашнее}$

задание No2 + 0,2*оценка за контрольную работу No1 + 0,2*оценка за контрольную работу No2 + 0,2*оценка за экзамен.

Все оценки выставляются по 10-балльной системе, способ округления всегда арифметический. При подсчете итоговой оценки накопленная берется с округлением вниз.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Примерный список вопросов на экзамене:

1. Cache-Oblivious бинарный поиск, принципы и устройство алгоритма.
2. FunnelSort – сортировка в модели Cache-Oblivious.
3. Структуры данных для поддержания квантилей в потоке чисел.
4. Поиск повторяющихся элементов в потоке чисел.
5. Нижние оценки на память структур данных, работающих над потоками данных.

Примерный список задач контрольной первого модуля:

1. Реализуйте сортировку в модели внешней памяти.
2. Реализуйте алгоритм List Ranking в модели внешней памяти.
3. Решите задачу подсчета размера поддеревьев в дереве в модели внешней памяти.

Примерный список задач контрольной второго модуля:

1. Проведите измерение латентности cache-ей процессора и оперативной памяти.
2. Реализуйте Cache-Oblivious бинарный поиск, проведите его сравнение со стандартной реализацией бинарного поиска.
3. Реализуйте один из алгоритмов приближенного подсчета количества элементов в потоке чисел (стоит выбрать MinCount - LogLog или HyperLogLog).

V. РЕСУРСЫ

V.1 Основная литература

1. U. Meyer, P. Sanders, and J. Sibeyn (eds.), "Algorithms for Memory Hierarchies", Springer-Verlag, 2003 (<http://www.win.tue.nl/~hermanh/teaching/2IL35/AMM/01-models-and-lower-bounds.pdf>), ([_http://www.win.tue.nl/~hermanh/teaching/2IL35/AMM/02-basic-data-structures.pdf](http://www.win.tue.nl/~hermanh/teaching/2IL35/AMM/02-basic-data-structures.pdf)), ([_http://www.win.tue.nl/~hermanh/teaching/2IL35/AMM/03-techniques.pdf](http://www.win.tue.nl/~hermanh/teaching/2IL35/AMM/03-techniques.pdf)), ([_http://www.win.tue.nl/~hermanh/teaching/2IL35/AMM/04-elementary-graph-algorithms.pdf](http://www.win.tue.nl/~hermanh/teaching/2IL35/AMM/04-elementary-graph-algorithms.pdf)) - очень неплохой обзор про алгоритмы по внешней памяти (основные примеры, оценки снизу, графы, структуры данных).
2. Jeff Erickson, "CS 473: Topics in Analysis of Algorithms (Fall 2003)" ([_http://compgeom.cs.uiuc.edu/~jeffe/teaching/473/](http://compgeom.cs.uiuc.edu/~jeffe/teaching/473/)) - записки лекций

Алгоритмы на графах

3. J. Abello, A. Buchsbaum, J. Westbrook, "A *Functional Approach to External Graph Algorithms*", Originally published in IBM Journal, Vol. 3, No.3. July, 1959
<http://citeseer.ist.psu.edu/viewdoc/download;jsessionid=CE9E1F1207053A2A612446D71F7D9AA3?doi=10.1.1.30.1027&rep=rep1&type=pdf>

V.2 Дополнительная литература

Cache-oblivious (экспериментальные результаты)

1. Brodal, Fagerberg, Jacob, "Cache-oblivious search trees via binary trees of small height" In Proc. 13th Ann. ACM-SIAM Symp. on Discrete, Algorithms, 2002.

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.6795&rep=rep1&type=pdf> -

производительность деревьев поиска, упакованных и индексированных в памяти различными способами.

V.3 Программное обеспечение

№ п/п	Наименование	Условия доступа
1.	Ubuntu 16 или выше	Свободно распространяемая система URL: http://releases.ubuntu.com/

V.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№ п/п	Наименование	Условия доступа
	<i>Профессиональные базы данных, информационно-справочные системы</i>	
1.	arxiv.org	URL: https://arxiv.org/
	<i>Интернет-ресурсы (электронные образовательные ресурсы)</i>	
1.	Открытое образование	URL: https://openedu.ru/

V.5 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные

программы, антивирусные программы);

– мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены персональными компьютерами (ноутбуками), с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.