

Министерство науки и образования РФ
Федеральное государственное автономное образовательное
учреждение высшего профессионального образования
«Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)»
(СПбГЭТУ «ЛЭТИ»)

Факультет компьютерных технологий и информатики

Кафедра вычислительной техники

**Пояснительная записка к курсовой работе по теме
«Измерение временной сложности алгоритма
в эксперименте на ЭВМ»
по дисциплине
«Алгоритмы и структуры данных»
Вариант 26**

Выполнил студент гр.9308:

Яловега Н.В.

Проверил:

Колинъко П.Г.

Оглавление

Введение.....	3
1. Задание.....	3
2. Формализация задания.....	3
3. Теоретические оценки временной сложности.....	4
4. Обработка результатов эксперимента.....	6
Вывод.....	8
Список используемых источников.....	9
Приложение.....	10

Введение

Исследование временной сложности алгоритмов над множествами последовательностями.

1. Задание

На основе программы, составленной по теме 3, можно выполнить статистический эксперимент по измерению фактической временной сложности алгоритма обработки данных.

2. Формализация задания

Программа из задания 3 дорабатывается таким образом, чтобы она генерировала множества мощностью, меняющейся, измеряла время выполнения цепочки операций над множествами и последовательностями и выводила результат в текстовый файл in.txt. Каждая строка этого файла должна по пару значений «размер входа — время» для каждого опыта. Затем эти данные обрабатываются, и по результатам обработки делается заключение о временной сложности алгоритма.

3. Теоретические оценки временной сложности

Вставка (insert)

В контейнере `set_seq` элемент вставляется и в дерево и в последовательность.

Вставка в дерево без указания места начала поиска может быть выполнена только за логарифмическое время $O(\log n)$, поскольку корректный поиск места вставки должен начинаться от корня дерева. Вставка за константное время $O(1)$ возможна, только если итератор вставки укажет для начала поиска места вставки на ключ, вставленный последним, однако такая вставка допустима только для упорядоченной последовательности ключей, которые вставляются в пустое дерево. В противном случае вставка будет вызывать хаос в структуре. Поэтому допустимо использовать вставку за константное время только для двуместных операциях с множествами по схеме слияния.

Вставка в последовательность осуществляется за константное время $O(1)$ (при условии, что контейнер не расширяется при добавлении в него элемента, если происходит перераспределение, то само перераспределение является линейным по всему размеру).

Поэтому в среднем временная сложность алгоритма вставки для операций над множествами по схемам слияния $O(1)$, а для произвольной вставки сложность составляет $O(\log n)$.

Объединение множеств (operator&=)

Для объединения по схеме слияния происходит не более $2*(N1+N2)-1$ сравнений, где $N1$ и $N2$ — размеры контейнеров, т.е. временная сложность алгоритма $O(n)$.

Разность множеств (operator-=)

Для разности по схеме слияния происходит не более $2*(N1+N2)-1$ сравнений, где $N1$ и $N2$ — размеры контейнеров, т.е. временная сложность алгоритма $O(n)$.

Симметрическая разность множеств (operator^=)

Для симметрической разности по схеме слияния происходит не более $2*(N1+N2)-1$ сравнений, где $N1$ и $N2$ — размеры контейнеров, т.е. временная сложность алгоритма $O(n)$.

Укорачивание (erase)

Временная сложность операции $O(n \log n)$, т.к. операция формирует результат одновременно со вставкой в дерево, причем вставка идет с поиском от корня, т.е. за логарифмическое время.

Исключение (excl)

Временная сложность операции линейная $O(n \log n)$, т.к. операция формирует результат одновременно со вставкой в дерево, причем вставка идет с поиском от корня, т.е. за логарифмическое время.

Замена (change).

Временная сложность операции линейная $O(n \log n)$, т.к операция формирует результат одновременно со вставкой в дерево, причем вставка идет с поиском от корня, т.е за логарифмическое время.

Таким образом, временная сложность цепочки операций $O(n \log n)$.

4. Обработка результатов эксперимента

Для эксперимента было выполнено 200 итераций. Мощность множеств генерировалась от 500 до 5500. Данные для эксперимента хранятся в файле in.txt

Результат работы программы RG41cb представлен на рисунке 1. Также он хранится в файле out.txt

201 data pairs received													Steps	Code
Dispersion	RMS	k	C	log N	N	N log N	N ²	N ³	N ⁴					
0.0525	0.229	1	0.294	0	0	0	0	0	0	0	0	0	201	++
0.00989	0.0995	2	-2.15	0.32	0	0	0	0	0	0	0	0	181	++
0.00222	0.0472	2	-0.126	0	0.000168	0	0	0	0	0	0	0	180	++
0.00134	0.0366	3	0.923	-0.162	0.000243	0	0	0	0	0	0	0	344	--
0.00133	0.0365	3	0.0558	0	-0.000616	8.92e-005	0	0	0	0	0	0	350	--
0.00134	0.0367	4	-0.0999	0.0286	-0.000759	0.000104	0	0	0	0	0	0	323	--
0.00148	0.0384	3	-0.0613	0	9.51e-005	0	1.43e-008	0	0	0	0	0	157	--
0.00146	0.0382	4	-0.0635	0	0.000257	-2.32e-005	2.19e-008	0	0	0	0	0	308	--
0.00131	0.0362	5	-0.126	0.0531	-0.00179	0.000235	-2.62e-008	0	0	0	0	0	749	**
0.00149	0.0386	4	-0.0634	0	9.6e-005	0	1.46e-008	-1.07e-013	0	0	0	0	326	**
0.00126	0.0355	5	-0.0949	0.0249	-9.49e-005	0	8.21e-008	-7.61e-012	0	0	0	0	592	**
0.00124	0.0351	5	-0.0658	0	0.00128	-0.000189	1.62e-007	-1.28e-011	0	0	0	0	521	**
0.00123	0.0351	6	-0.0297	-0.0333	0.00308	-0.000434	2.6e-007	-1.9e-011	0	0	0	0	963	**
0.00121	0.0349	5	-0.0648	0	0.000194	0	-1.01e-007	3.97e-011	-4.21e-015	0	0	0	509	**
0.00122	0.0349	6	-0.0661	0	0.000648	-7.78e-005	1.79e-009	1.99e-011	-2.66e-015	0	0	0	649	**
0.00122	0.0349	7	-0.133	0.0611	-0.00358	0.000534	-4.09e-007	7.8e-011	-6.44e-015	0	0	0	1485	**

Результаты расчёта отношений выборочных дисперсий

Вспомогательная таблица. Определение подходящего уравнения регрессии																
Отношение дисперсий																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1,00	0,19	0,04	0,03	0,03	0,03	0,03	0,03	0,02	0,03	0,02	0,02	0,02	0,02	0,02	0,02
2	5,30	1,00	0,22	0,14	0,13	0,14	0,15	0,15	0,13	0,15	0,13	0,12	0,12	0,12	0,12	0,12
3	23,59	4,45	1,00	0,60	0,60	0,60	0,66	0,66	0,59	0,67	0,57	0,56	0,55	0,55	0,55	0,55
4	39,11	7,38	1,66	1,00	0,99	1,00	1,10	1,09	0,98	1,11	0,94	0,92	0,92	0,91	0,91	0,91
5	39,43	7,44	1,67	1,01	1,00	1,01	1,11	1,10	0,98	1,12	0,95	0,93	0,93	0,91	0,92	0,91
6	39,04	7,36	1,65	1,00	0,99	1,00	1,10	1,09	0,97	1,11	0,94	0,92	0,92	0,90	0,91	0,91
7	35,49	6,69	1,50	0,91	0,90	0,91	1,00	0,99	0,89	1,01	0,85	0,84	0,83	0,82	0,82	0,82
8	35,85	6,76	1,52	0,92	0,91	0,92	1,01	1,00	0,89	1,02	0,86	0,84	0,84	0,83	0,83	0,83
9	40,07	7,56	1,70	1,02	1,02	1,03	1,13	1,12	1,00	1,14	0,96	0,94	0,94	0,93	0,93	0,93
10	35,11	6,62	1,49	0,90	0,89	0,90	0,99	0,98	0,88	1,00	0,84	0,83	0,82	0,81	0,82	0,81
11	41,72	7,87	1,77	1,07	1,06	1,07	1,18	1,16	1,04	1,19	1,00	0,98	0,98	0,97	0,97	0,97
12	42,47	8,01	1,80	1,09	1,08	1,09	1,20	1,18	1,06	1,21	1,02	1,00	1,00	0,98	0,99	0,99
13	42,62	8,04	1,81	1,09	1,08	1,09	1,20	1,19	1,06	1,21	1,02	1,00	1,00	0,99	0,99	0,99
14	43,18	8,14	1,83	1,10	1,10	1,11	1,22	1,20	1,08	1,23	1,03	1,02	1,01	1,00	1,00	1,00
15	43,07	8,12	1,83	1,10	1,09	1,10	1,21	1,20	1,07	1,23	1,03	1,01	1,01	1,00	1,00	1,00
16	43,11	8,13	1,83	1,10	1,09	1,10	1,21	1,20	1,08	1,23	1,03	1,02	1,01	1,00	1,00	1,00

На основании критерия Фишера была выбрана линейная регрессия. Отношения первой к остальным дисперсиям больше 43, следовательно, сложность не константная.

Отношения второй к остальным дисперсиям больше 8, следовательно, сложность не логарифмическая.

Отношения же третьей к остальным дисперсиям превышает 1.8.

Отношения четверной к остальным дисперсиям не превышает 1.26, из чего можно сделать вывод, что уравнение регрессии — четвертое. Усложнение моделей регрессии после четвертой — незначимо.

4. Линейная регрессия - 2

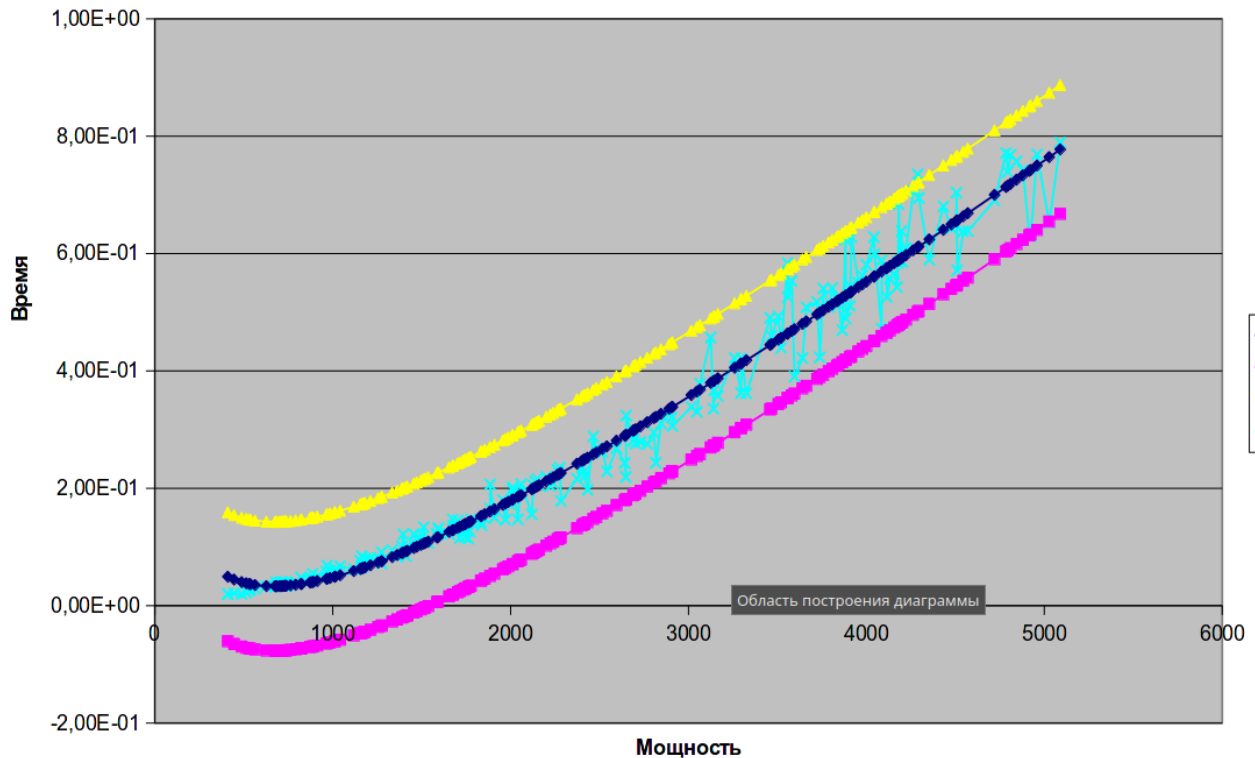


График подтверждает достоверность результатов обработки. Измеренные значения усреднены кривой регрессии, располагаясь примерно поровну выше и ниже её. Практически все отсчёты находятся в доверительном интервале $\pm 3 \sigma$, что соответствует допустимой ошибке 0,03% (нормальное распределение).

Для более сложных моделей наблюдается такая же картина, с примерно тем же размером доверительного интервала, что и означает бессмысленность усложнения модели.

Вывод

Ожидаемая временная сложность цепочки операций была $O(n \log n)$. Исходя из практической оценки, можно сделать вывод, что временная сложность линейная $O(n)$. Некоторые возможные причины получения недостоверных результатов были приняты во внимание:

Проводилось 200 опытов, что более чем достаточно. Для оценки использовалась фактическая средняя мощность множеств, т.е. результаты эксперимента не получены со смещением. Также нет выбросов (данных, которые сильно отличаются от средних значений, отсчётов вне доверительного интервала). При просмотре вывода программы в отладочном режиме не было замечено некорректных примеров.

Возможная причина получения такого результата:

Для учета повторяющихся значений в последовательности было сделано отдельное поле в структуре узла. Поэтому вставка в дерево уже существующего ключа будет меньше чем $O(\log n)$. Однако, максимальное значение ключа не константное (как было предложено в тестовой программе), а зависит от мощности генерируемого множества (в три раза больше). При таком подходе для множеств большой мощности не должно быть избытка дубликатов.

При выполнении курсовой работы были получены практические навыки в написании программ на языке C++ с вычислением практической временной сложности программы. В результате работы были получены данные, которые опровергли теоретическую оценку: по окончании 3-ей лабораторной работы был выдвинут тезис о сложности алгоритма $O(n \log n)$, однако в ходе курсовой выяснили, что временная сложность - линейная.

Список используемых источников

1. Колинко П.Г. Пользовательские структуры данных Часть 2: Методические указания к практическим занятиям на ПЭВМ и курсовому проектированию. / Методические указания по дисциплине «Алгоритмы и структуры данных» - Санкт-Петербург: СПбГЭТУ «ЛЭТИ», 2020.
2. Скиена С. С Алгоритмы. Руководство по разработке. — 2-е изд.: Пер. с англ. — СПб.: БХВ-Петербург, 2018. — 720 с.: ил

Приложение

avl_tree.h — реализация АВЛ-дерева

main.cpp — программа для эксперимента

in.txt — данные, полученные в ходе эксперимента

out.txt — данные, полученные программой RG41cb

table.xls — таблица