Тема: Идеально сбалансированное дерево поиска (ИСДП)

Цель работы: Изучение процесса программного построения ИСДП.

1. Постановка задачи

1. Написать подпрограммы для вычисления характеристик двоичного дерева, которые определяют

• размер дерева;

• высоту дерева;

• среднюю высоту дерева;

• контрольную сумму данных в вершинах дерева;

и проверить их работу на конкретном примере.

2. Запрограммировать обход двоичного дерева слева направо и вывести на экран получившуюся последовательность данных.

3. Разработать подпрограмму поиска вершины с заданным ключом в двоичном дереве поиска.

4. Разработать подпрограмму построения идеально сбалансированного дерева поиска (ИСДП) для массива случайных чисел, а также логическую функцию для определения является ли данное двоичное дерево деревом поиска. Построить ИСДП из 100, 200,…, 500 вершин (данные в вершинах произвольные, но все различные). Распечатать обход дерева слева направо. Для построенных деревьев вычислить размер, контрольную сумму, высоту и среднюю высоту, используя разработанные функции. Заполнить таблицу и проанализировать полученные результаты.

5. Разработать подпрограмму построения случайного дерева поиска (СДП). Построить СДП из 100, 200,…, 500 вершин (данные в вершинах произвольные, но все различные). Распечатать обход дерева слева направо. Для построенного дерева вычислить размер, контрольную сумму, высоту и среднюю высоту, сравнить их с аналогичными характеристиками ИСДП. ИСДП необходимо строить для той же последовательности данных, что и СДП. Заполнить таблицу и проанализировать полученные результаты.

2. Описание используемых алгоритмов

Определим двоичное дерево следующим образом:

1. Отдельная вершина V является двоичным деревом.

2. Двоичное дерево – это вершина V, соединенная с (возможно пустыми) левым (ТL) и правым (ТR) двоичными деревьями.

Каждая вершина дерева может содержать какую-либо информацию. Выделенная вершина дерева называется корнем. Концевые вершины дерева, не имеющие поддеревьев, называются листьями. Дуги ориентированы по направлению от корня к листьям. Путь от корня к листу называется ветвью. Под длиной ветви будем понимать число входящих в неё вершин. Высота дерева h определяется как число вершин в самой длинной ветви дерева. Размер дерева – число входящих в него вершин. Средней высотой дерева называется усредненная по количеству вершин в дереве сумма длин путей от корня к каждой вершине.

Обходы легко программируются с помощью рекурсивных процедур.

Алгоритм на псевдокоде

Обход слева направо(p: pVertex)

 IF (p ≠ NIL)

 Обход слева направо (p→Left)

 Печать (p→Data)

 Обход слева направо (p→Right)

 FI

Приведем алгоритмы для вычисления различных характеристик двоичных деревьев.

1) Определение размера дерева

Алгоритм на псевдокоде

Размер(p: pVertex)

 IF (p = NIL) Размер := 0

 ELSE Размер := 1 + Размер (p→Left) + Размер (p→Right)

 FI

2) Определение высоты дерева

Алгоритм на псевдокоде

Высота(p: pVertex)

 IF (p = NIL) Высота := 0

 ELSE Высота := 1+ max(Высота (p→Left), Высота(p→Right))

 FI

3) Определение средней высоты дерева

Для определения средней высоты дерева понадобится функция вычисления суммы длин путей от корня до каждой вершины на L-том уровне.

Алгоритм на псевдокоде

СДП (p: pVertex; L: уровень вершины)

 IF (p = NIL) СДП:= 0

 ELSE СДП:= L + СДП(p → Left, L+1) + СДП(p → Right, L+1)

 FI

Тогда средняя высота вычисляется следующим образом

hcp := СДП(Root, 1)/ Размер(Root)

4) Определение контрольной суммы для дерева

Алгоритм на псевдокоде

Сумма (p: pVertex)

 IF (p = NIL) Сумма := 0

 ELSE Сумма:= p→Data + Сумма(p→Left) + Сумма(p→Right )

 FI

Двоичное дерево называется деревом поиска, если ключ в каждой его вершине больше ключа любой вершины в левом поддереве и меньше ключа любой вершины в правом поддереве.

Двоичное дерево называется идеально сбалансированным (ИСД), если для каждой его вершины размеры левого и правого поддеревьев отличаются не более чем на 1.

Отметим некоторые свойства идеально сбалансированного дерева.

Свойство 1. Высота ИСД с n вершинами минимальна и равна

*hИСД(n) = hmin (n) =*⎡log*(n+1)*⎤.

.

Свойство 2. Если дерево идеально сбалансировано, то все его поддеревья также идеально сбалансированы.

Задача построения идеально сбалансированного дерева поиска из элементов массива А = (а1, а2, … , аn) решается в два этапа:

1. Сортировка массива А.

2. В качестве корня дерева возьмем средний элемент отсортированного массива, из меньших элементов массива строим левое поддерево, из больших – правое поддерево. Далее процесс построения продолжается до тех пор, пока левое или правое поддерево не станет пустым.

Приведем на псевдокоде алгоритм построения ИСДП. Функция ИСДП (L, R) возвращает указатель на построенное дерево, где L, R – левая и правая границы той части массива, из элементов которой строится дерево.

Алгоритм на псевдокоде

ИСДП (L, R)

IF (L > R) ИСДП:=NIL

ELSE m : = [(L+R) / 2]

<Выделяем память для p>

p → Data : = A[m]

p → Left : = ИСДП ( L, m-1)

p → Right : = ИСДП ( m+1, R)

ИСДП:= p

FI

Для идеально сбалансированного дерева Сmax = 2⎡log(n+1)⎤. Если считать, что поиск любой вершины происходит одинаково часто, то ИСДП обеспечивает минимальное среднее время поиска. К существенным недостаткам ИСДП можно отнести то, что при добавлении нового элемента к множеству данных необходимо строить заново ИСДП.

3. Результат работы программы

Фрагменты работы программы показаны на рисунках.





Результаты замеров для ИСДП выводятся в файл Balanced.txt.

Результаты замеров для ИСДП выводятся в файл Random.txt.

4. Анализ и сравнение полученных результатов с теоретическими оценками

|  |  |  |
| --- | --- | --- |
| Размер дерева | СДП | ИСДП |
| Контр.сумма | Высота фактическая | Теор. оценки для сред. высоты | Контр.сумма | Высота фактическая | Теор. оценки для сред. высоты |
| 100 |  9652 |  12 |  5.95 | 9652 | 7 |  4.80 |
| 200 |  40815 |  16 |  8.37 | 40815 | 8 |  5.76 |
| 300 |  88492 |  17 |  8.54 | 88492 | 9 |  6.33 |
| 400 |  163252 |  21 |  9.29 | 163252 | 9 |  6.75 |
| 500 |  246355 |  18 |  9.49 | 246355 | 9 |  7.00 |

Как видно из результатов замеров, для одного и того же количества элементов СДП имеет высоту больше, чем ИСДП.