Тема 2. СТРУКТУРЫ ДАННЫХ

2.5. Деревья

2.5.5. Расширенные бинарные деревья

Любое непустое бинарное дерево можно дополнить фиктивными вершинами-листьями так, чтобы все вершины исходного дерева стали внутренними и имели точно по два сына (рис. 2.17). Построенное таким образом дерево называется *расширенным бинарным деревом*, а добавленные фиктивные листья (изображены квадратами) – *внешними вершинами* (или *внешними узлами* при узловом представлении дерева).

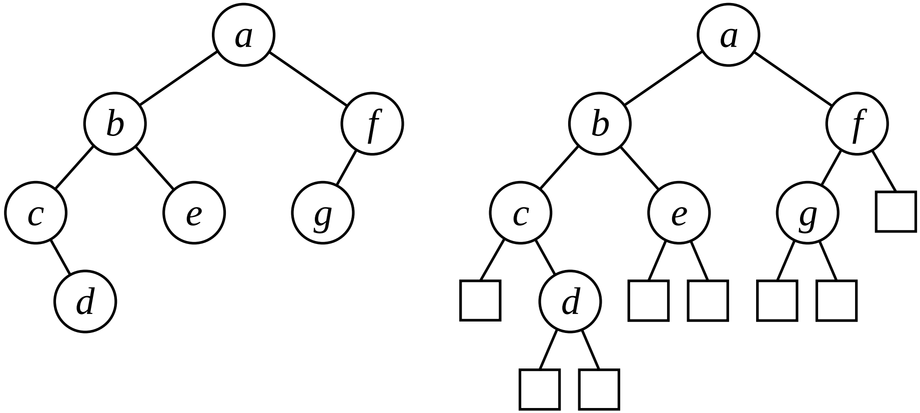

Рис. 2.17. Бинарное дерево и соответствующее ему расширенное дерево

Деревья используются не только как структуры данных для представления множеств или иерархических структур, но и для анализа определенных алгоритмов. При анализе алгоритмов их поведение можно представить в виде так называемых *деревьев решений* (дерево поиска, дерево сортировки, дерево игры и т. п.). Поэтому возникает необходимость в количественных измерениях различных характеристик деревьев.

В расширенном бинарном дереве с n внутренними вершинами всегда существует $n + 1$ внешних вершин. *Длина внешних путей* $E(T)$ расширенного бинарного дерева T с n внутренними вершинами определяется как сумма уровней всех внешних вершин. *Длина внутренних путей* $I(T)$ есть сумма уровней всех внутренних вершин. У расширенного бинарного дерева (рис. 2.17) длина внешних путей $E(T) = 25$, а длина внутренних путей $I(T) = 11$.

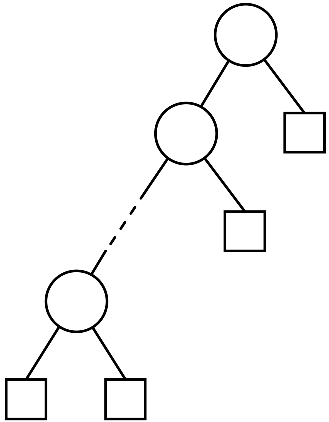

Рис. 2.18. Расширенное бинарное дерево с максимальными длинами путей

Среднее расстояние (т. е. средняя длина) определяется делением длины путей на соответствующее число вершин, т. е. среднее расстояние до внешней вершины равно $E(T)/(n + 1)$, а среднее расстояние до внутренней вершины равно $I(T)/n$.

Существует соотношение между длинами внешних и внутренних путей. Для расширенного бинарного дерева

$$E(T) = I(T) + 2n. \qquad (2.1)$$

С точки зрения анализа алгоритмов интересен диапазон значений длин путей. Максимальную длину путей имеют бинарные деревья, вырожденные до связного списка (рис. 2.18). В этом случае расширенное бинарное дерево T с n внутренними вершинами имеет длины

$$I_{\max}(T) = \sum_{i=0}^{n-1} i = \frac{1}{2} n(n-1) \quad \text{и} \quad E_{\max}(T) = \frac{1}{2} n(n+3).$$

Если l_1, l_2, …, l_{n+1} – уровни $n + 1$ внешних вершин в расширенном бинарном дереве с n внутренними вершинами, то

$$\sum_{i=1}^{n+1} \frac{1}{2^{l_i}} = 1. \qquad (2.2)$$

Это соотношение легко доказывается методом индукций.

Минимальную длину путей имеют полностью сбалансированные деревья. Бинарное дерево называется *полностью сбалансированным*, если у расширенного дерева все внешние вершины находятся на уровнях l и $l + 1$, где l – некоторое целое число. Если все внешние вершины находятся на одном уровне l, то такое дерево называется *полным*.

Определим минимальную длину внешних путей для полностью сбалансированного бинарного дерева с n внутренними и $n + 1$ внешними вершинами. Пусть на уровне l находится k внешних вершин, тогда на уровне $l + 1$ будет $n + 1 – k$ внешних вершин, причем $1 \le k \le n + 1$ (при $k = n + 1$ все внешние вершины находятся на уровне l). Из соотношения (2.2) следует, что

$$\frac{k}{2^l}+\frac{n+1-k}{2^{l+1}}=1.$$

Отсюда

$$k=2^{l+1}-n-1. \qquad (2.3)$$

Поскольку $k \le n + 1$, то $2^l \le n + 1$. Следовательно,

$$l=\lfloor \log(n+1) \rfloor. \qquad (2.4)$$

Объединение (2.3) и (2.4) дает

$$k=2^{\lfloor \log(n+1) \rfloor+1}-n-1.$$

Таким образом, минимальная длина внешних путей

$$E_{\min}(T) = l\,k + (l+1)(n+1-k) =$$
$$= (n+1)\lfloor \log(n+1) \rfloor + 2(n+1) - 2^{\lfloor \log(n+1) \rfloor + 1}.$$

Полученную формулу можно представить в виде

$$E_{\min}(T) = (n+1)\log(n+1) + (n+1)(2 - \theta - 2^{1-\theta}), \quad (2.5)$$

где

$$\theta = \log(n+1) - \lfloor \log(n+1) \rfloor,\ 0 \le \theta < 1.$$

Следует отметить, что $\theta = 0$, если все внешние вершины находятся на уровне $l = \log(n+1)$, т. е. для полных бинарных деревьев число внутренних вершин $n = 2^l - 1$, а число внешних вершин является степенью числа 2.

Очевидно, что высота полностью сбалансированного расширенного бинарного дерева

$$h(T) = \lceil \log(n+1) \rceil.$$

Формулу для определения минимальной длины внутренних путей можно получить с помощью соотношения (2.1).

Понятие расширенного дерева легко обобщить на m-арные деревья, в которых все внутренние вершины имеют степень m. В расширенном m-арном дереве с n внутренними вершинами имеется $(m-1)n+1$ внешних вершин. Аналогом соотношения (2.1) будет соотношение

$$E(T) = (m-1)\, I(T) + m\, n.$$

Минимальная длина внешних путей равна

$$(n(m-1)+1)l - \frac{m^{l+1}-m}{m-1} + mn, \text{ где } l = \lfloor \log_m (n(m-1)+1) \rfloor.$$