Тема: «Бинарное дерево»

Введение

При решении многих задач (реализация СУБД, разработка компиляторов и т.п.) возникает необходимость в применении высокопроизводительных алгоритмов доступа к данным. Как известно, для ускорения доступа к информации применяются индексы. Есть много типов индексов: битовые, кластерные хеш-индексы, но, наверное, самым популярными и распространенными являются индексы на основе бинарных деревьев.

Общие понятия

Массивы и связанные списки определяют коллекции объектов, доступ к которым осуществляется последовательно. Такие структуры данных называют линейными (linear) списками, поскольку они имеют уникальные первый и последний элементы и у каждого внутреннего элемента есть только один наследник. Линейный список является абстракцией, позволяющей манипулировать данными, представляемыми различным образом — массивами, стеками, очередями и связанными списками.

Во многих приложениях обнаруживается нелинейный порядок объектов, где элементы могут иметь нескольких наследников. Например, в фамильном дереве родитель может иметь нескольких потомков (детей). На рис. 1 показаны три поколения семьи. Такое упорядочение называют иерархическим.



Рассмотрим нелинейную структуру, называемую деревом (tree), которая состоит из узлов и ветвей и имеет направление от корня к внешним узлам, называемым листьями.

Древовидная структура характеризуется множеством узлов (nodes), происходящих от единственного начального узла, называемого корнем (root). На рис. 2 корнем является узел А. В терминах генеалогического дерева узел можно считать родителем (parent), указывающим на 0, 1 или более узлов, называемых сыновьями (children). Например, узел В является родителем сыновей E и F. Родитель узла H - узел D. Дерево может представлять несколько поколений семьи. Сыновья узла и сыновья их сыновей называются потомками (descendants), а родители и прародители – предками (ancestors) этого узла. Например, узлы E, F, I, J – потомки узла B. Каждый некорневой узел имеет только одного родителя, и каждый родитель имеет 0 или более сыновей. Узел, не имеющий детей (E, G, H, I, J), называется листом (leaf).

Каждый узел дерева является корнем поддерева (subtree), которое определяется данным узлом и всеми потомками этого узла. Узел F есть корень поддерева, содержащего узлы F, I и J. Узел G является корнем поддерева без потомков. Это определение позволяет говорить, что узел A есть корень поддерева, которое само оказывается деревом.



Переход от родительского узла к дочернему и к другим потомкам осуществляется вдоль пути (path). Например, на рис. 2 путь от корня A к узлу F проходит от A к B и от B к F. Тот факт, что каждый некорневой узел имеет единственного родителя, гарантирует, что существует единственный путь из любого узла к его потомкам. Длина пути от корня к этому узлу есть **уровень узла**. **Уровень корня равен 0**. Каждый сын корня является узлом 1-го уровня, следующее поколение – узлами 2-го уровня и т.д. Например, на рис. 3 узел F является узлом 2-го уровня с длиной пути 2.

**Глубина (depth) дерева есть его максимальный уровень**. Понятие глубины также может быть описано в терминах пути. Глубина дерева есть длина самого длинного пути от корня до узла. На рис. 3 глубина дерева равна 3.



Бинарные деревья

Мы сосредоточимся на ограниченном классе деревьев, где каждый родитель имеет не более двух сыновей (рис. 4). Такие бинарные деревья (binary trees) имеют унифицированную структуру (данные передаются от одного узла к другому), допускающую разнообразные алгоритмы прохождения и эффективный доступ к элементам.

У каждого узла бинарного дерева может быть 0, 1 или 2 сына. По отношению к узлу слева будем употреблять термин левый сын (left child), а по отношению к узлу справа – правый сын (right child). Наименования "левый" и "правый" относятся к графическому представлению дерева. Бинарное дерево является рекурсивной структурой. Каждый узел – это корень своего собственного поддерева. У него есть сыновья, которые сами являются корнями деревьев, называемых левым и правым поддеревьями соответственно. Таким образом, процедуры обработки деревьев рекурсивны.

Бинарное дерево — это конечное множество элементов, связанных с двумя разными бинарными деревьями — правым и левым поддеревьями.



Рис. 4. Структура бинарного дерева



Необходимо выделить следующие термины:

— узел (вершина) — это каждый элемент бинарного дерева;

— ветви — связи между узлами;

— глубина (высота) — наибольший уровень какого-нибудь элемента;

— лист (терминальный узел) — термин применяется, если элемент не имеет потомков;

— внутренние узлы — узлы ветвления;

— степень внутреннего узла — число его потомков (наибольшая степень всех существующих узлов — это степень всего бинарного дерева);

— длина пути к x — количество ветвей, которые необходимо пройти от корня до текущего узла x. Длина пути самого корня равна нулю, а узел на уровне i обладает длиной пути, которая равна i.