МГТУ им. Н.Э. Баумана
ТЕХНОПАРК
Mail.Ru Group

АЛГОРИТМЫ И СТРУКТУРЫ ДАННЫХ

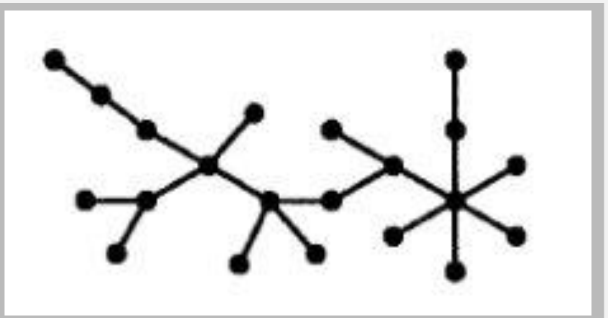
Лекция 6

Мацкевич С.Е.

План лекции 6 «Деревья»

1. Определения, примеры деревьев
2. Представление в памяти
3. Обходы дерева в глубину, в ширину
4. Двоичные деревья поиска
5. Декартовы деревья
6. АВЛ-деревья
7. АТД «Ассоциативный массив»

Определения деревьев



<u>Определение 1.</u> Дерево (свободное) – непустая коллекция **вершин** и **ребер**, удовлетворяющих **определяющему свойству дерева**.

Вершина (узел) – простой объект, который может содержать некоторую информацию.

Ребро – связь между двумя вершинами.

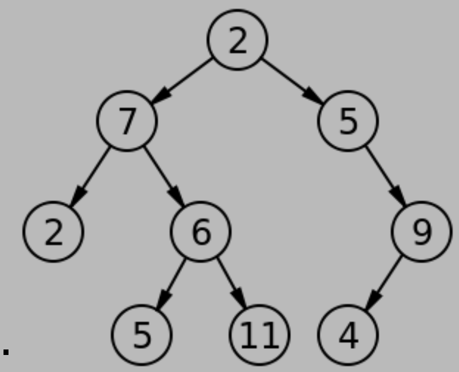
Путь в дереве – список отдельных вершин, в котором следующие друг за другом вершины соединяются ребрами дерева.

Определяющее свойство дерева – существование только одного пути, соединяющего любые два узла.

<u>Определение 2</u> (равносильно первому). Дерево (свободное) – неориентированный связный граф без циклов.

Определения деревьев

<u>**Определение 3.**</u> **Дерево с корнем** — дерево, в котором один узел выделен и назначен «корнем» дерева.



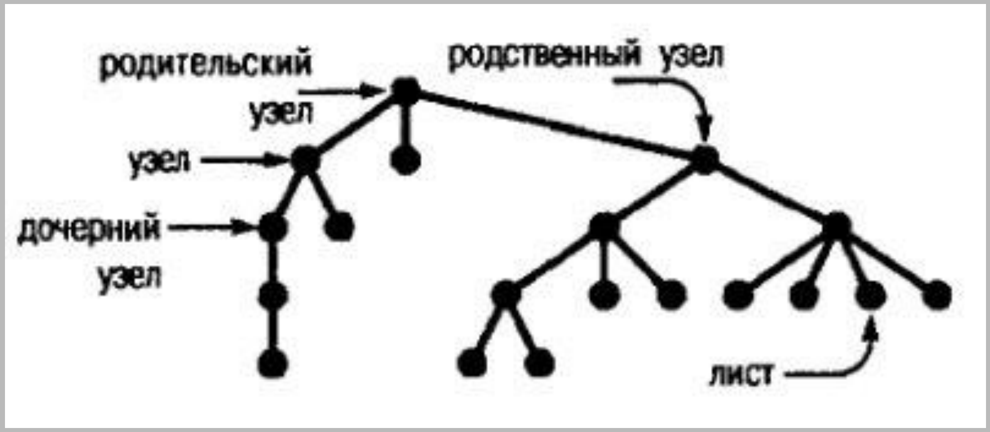
Существует только один путь между корнем и каждым из других узлов дерева.

<u>**Определение 4.**</u> **Высота (глубина)** дерева с корнем — количество вершин в самом длинном пути от корня.

Обычно дерево с корнем рисуют с корнем, расположенным сверху. Узел y располагается под узлом x (а x располагается над y), если x располагается на пути от y к корню.

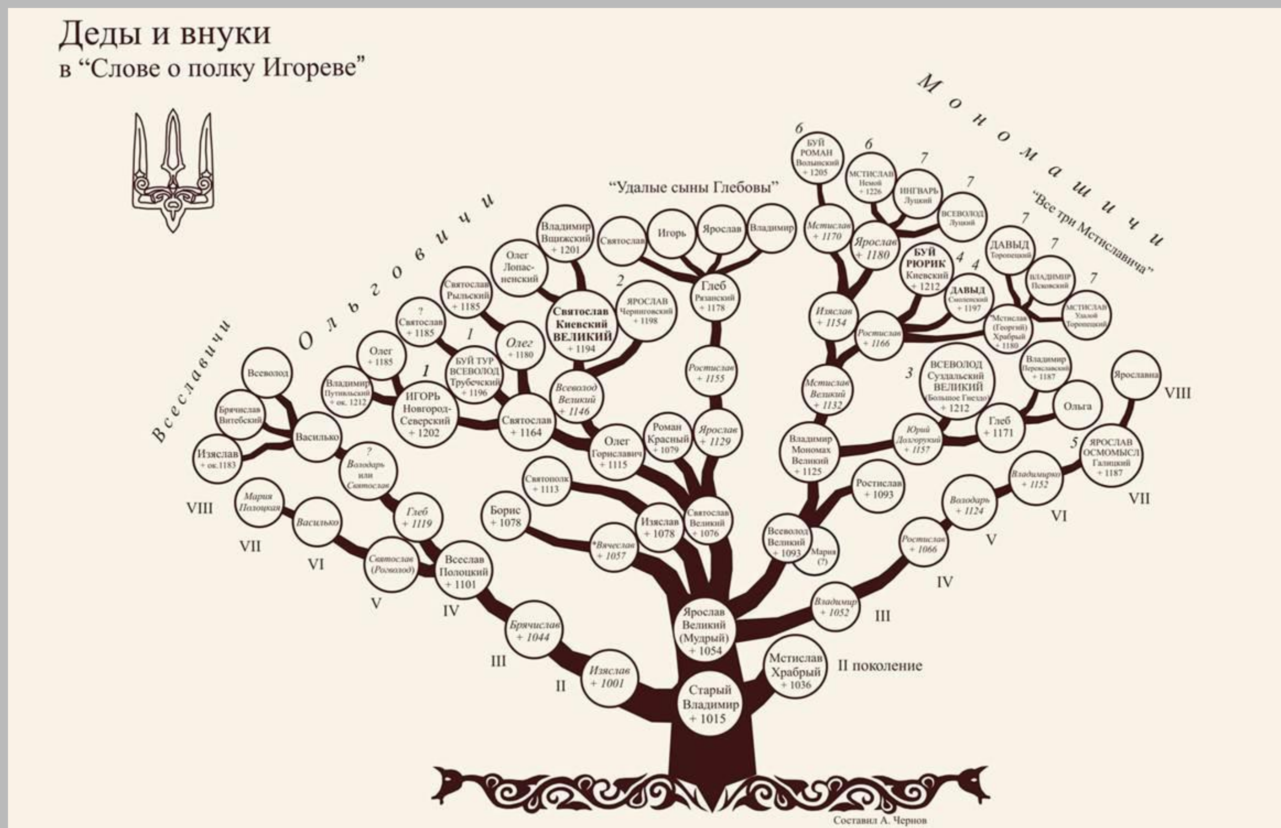
Определения деревьев

Определение 5. Каждый узел (за исключением корня) имеет только один узел, расположенный над ним. Такой узел называется **родительским.**
Узлы, расположенные непосредственно под данным узлом, называются его **дочерними** узлами.
Узлы, не имеющие дочерних узлов называются **листьями**.



Примеры деревьев

Генеалогическое дерево



Примеры деревьев

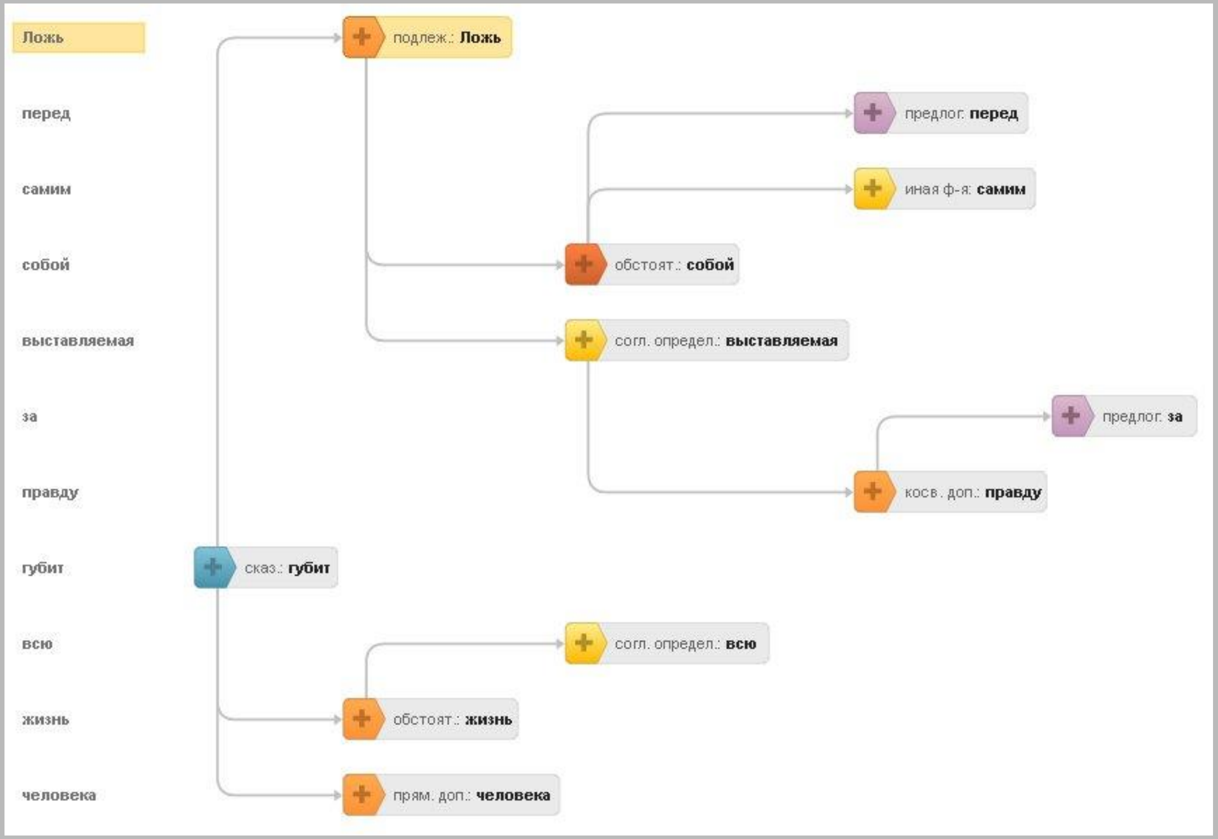
Организация турнира.

Примеры деревьев

Орг. структура компании.

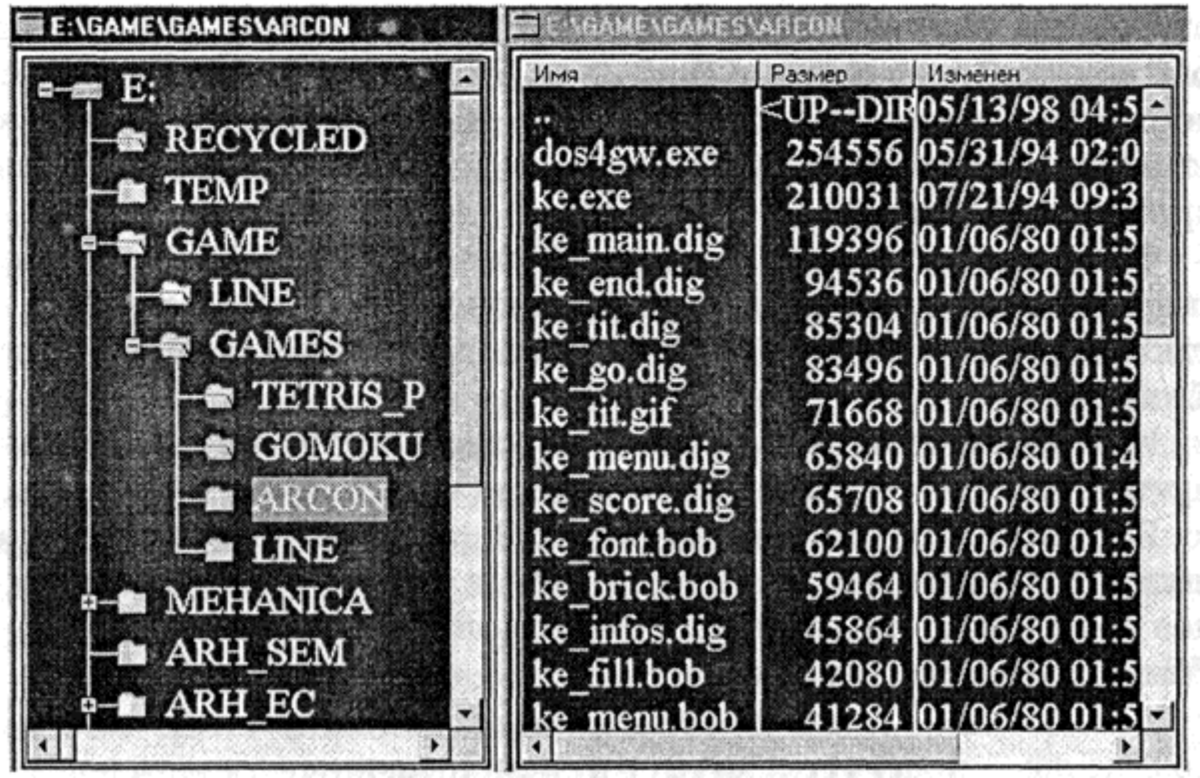
Примеры деревьев

Синтаксический или семантический разбор предложения.



Примеры деревьев

Файловая система.



Число вершин и ребер

<u>Утверждение 1.</u> Любое дерево (с корнем) содержит листовую вершину.

<u>Доказательство.</u> Самая глубокая вершина является листовой.

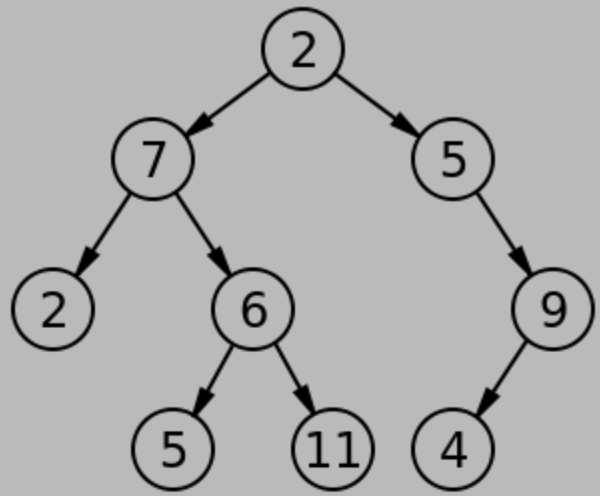
<u>Утверждение 2.</u> Дерево, состоящее из N вершин, содержит $N - 1$ ребро.

<u>Доказательство.</u> По индукции.
База индукции. $N = 1$. Одна вершина, ноль ребер.
Шаг индукции. Пусть дерево состоит из $N + 1$ вершины. Найдем листовую вершину. Эта вершина содержит ровно 1 ребро. Дерево без этой вершины содержит N вершин, а по предположению индукции $N - 1$ ребро. Следовательно, исходное дерево содержит N ребер, ч.т.д.

Виды деревьев

<u>Определение 6а.</u> Двоичное (бинарное) дерево —
это дерево, в котором степени вершин не превосходят 3.

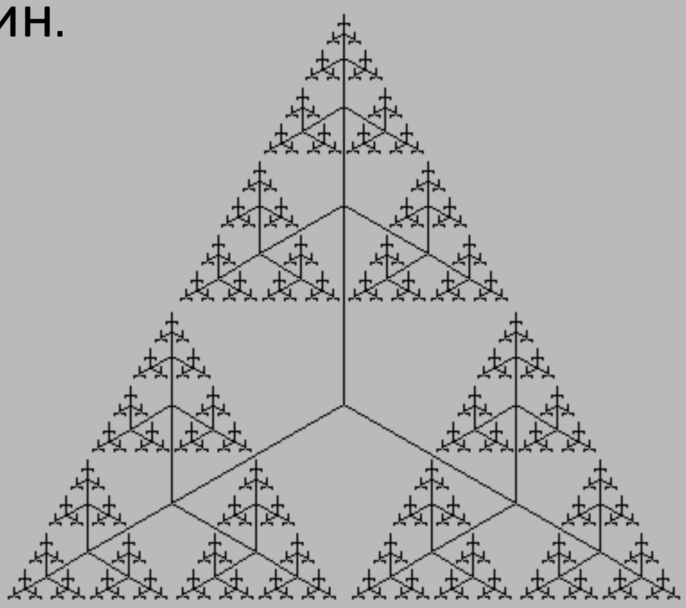
<u>Определение 6б.</u> Двоичное (бинарное) дерево с корнем — это дерево, в котором каждая вершина имеет не более двух дочерних вершин.



Виды деревьев

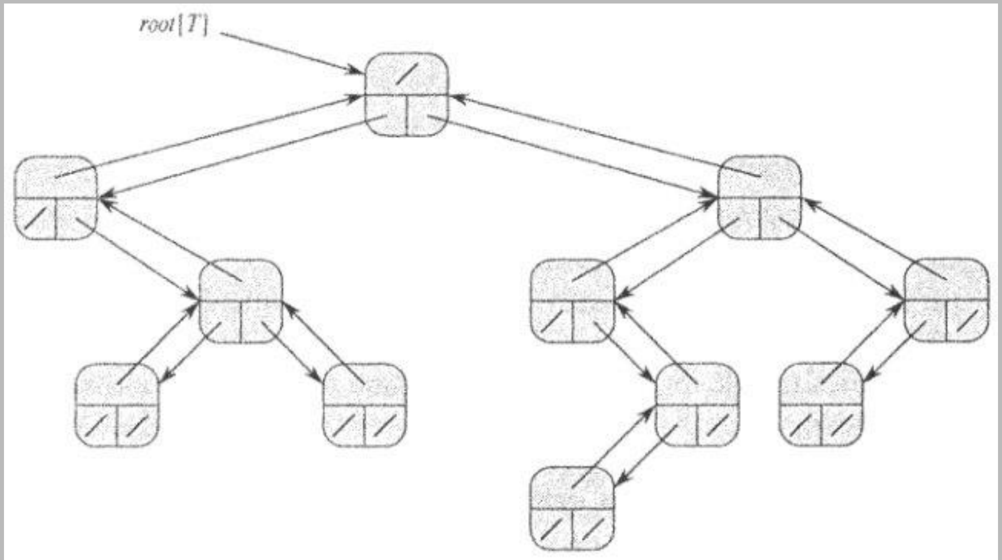
Определение 7а. N-арное дерево — это дерево, в котором степени вершин не превосходят N + 1.

Определение 7б. N-арное дерево с корнем — это дерево, в котором каждая вершина имеет не более N дочерних вершин.



Структуры данных

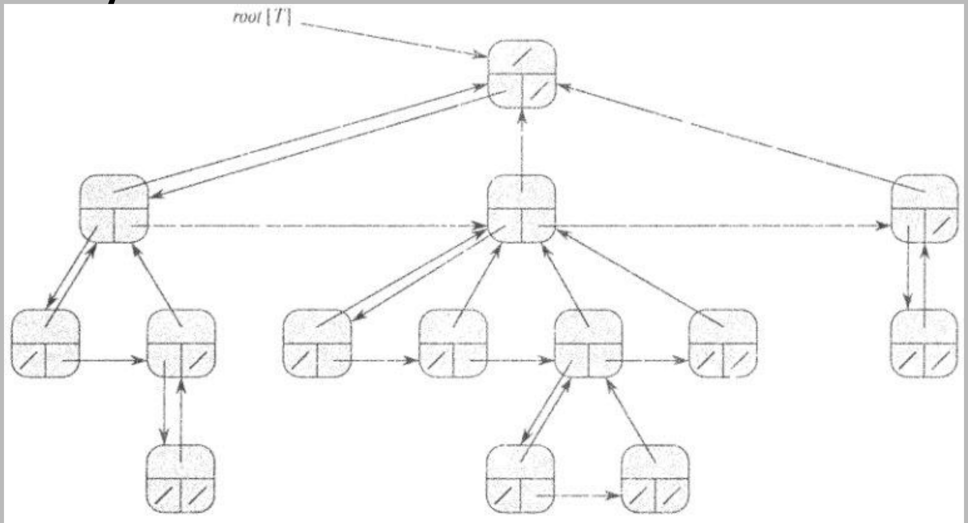
<u>Определение 8.</u> СД «Двоичное дерево» — представление двоичного дерева с корнем.
Узел – структура, содержащая данные и указатели на левый и правый дочерний узел. Также может содержать указатель на родительский узел.



Структуры данных

<u>Определение 9.</u> СД «N-арное дерево» — представление N-арного дерева с корнем.

Узел – структура, содержащая данные, указатель на следующий родственный узел и указатель на первый дочерний узел. Также может содержать указатель на родительский узел.



Структуры данных

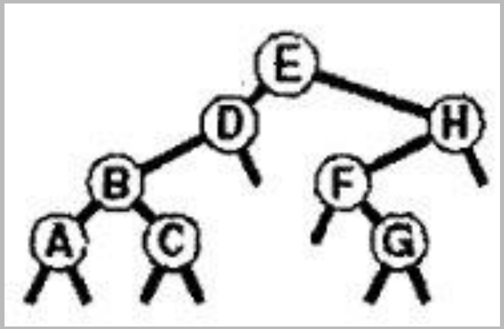
```
// Узел двоичного дерева с данными типа int.
struct CBinaryNode {
    int Data;
    CBinaryNode* Left; // NULL, если нет.
    CBinaryNode* Right; // NULL, если нет.
    CBinaryNode* Parent; // NULL, если корень.
};

// Узел дерева с произвольным ветвлением.
struct CTreeNode {
    int Data;
    CTreeNode* Next; // NULL, если нет следующих.
    CTreeNode* First; // NULL, если нет дочерних.
    CTreeNode* Parent; // NULL, если корень.
};
```

Обход дерева в глубину

Определение 10. Пошаговый перебор элементов дерева по связям между узлами-предками и узлами-потомками называется **обходом дерева**.

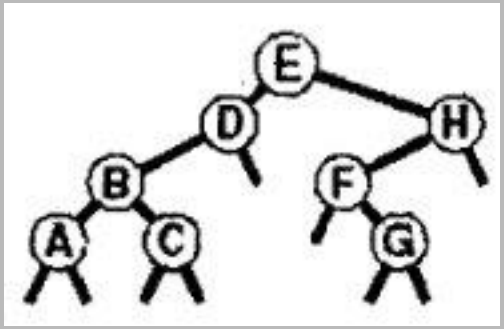
Определение 11. Обходом двоичного дерева в глубину (DFS) называется процедура, выполняющая в некотором заданном порядке следующие действия с поддеревом:
* просмотр (обработка) узла-корня поддерева,
* рекурсивный обход левого поддерева,
* рекурсивный обход правого поддерева.
DFS – Depth First Search.

Обход дерева в глубину



- **Прямой обход (сверху вниз, pre-order).** Вначале обрабатывается узел, затем посещается левое и правые поддеревья. Порядок обработки узлов дерева на рисунке: E, D, B, A, C, H, F, G.
- **Обратный обход (снизу вверх, post-order).** Вначале посещаются левое и правое поддеревья, а затем обрабатывается узел. Порядок обработки узлов дерева на рисунке: A, C, B, D, G, F, H, E.

Обход дерева в глубину



- **Поперечный обход (слева направо, in-order)**. Вначале посещается левое поддерево, затем узел и правое поддерево.
  Порядок обработки узлов дерева на рисунке:
  A, B, C, D, E, F, G, H.

Обход дерева в глубину

```
// Обратный обход в глубину.
void TraverseDFS( CBinaryNode* node )
{
    if( node == 0 )
        return;
    TraverseDFS( node->Left );
    TraverseDFS( node->Right );
    visit( node );
};
```

Обход дерева в глубину

Задача. Вычислить количество вершин в дереве.
Решение. Обойти дерево в глубину в обратном порядке. После обработки левого и правого поддеревьев вычисляется число вершин в текущем поддереве.

Реализация:

```
// Возвращает количество элементов в поддереве.
int Count( CBinaryNode* node )
{
    if( node == 0 )
        return 0;
    return Count( node->Left ) + Count( node->Right ) + 1;
};
```

Обход дерева в глубину

- **Обход в глубину** не начинает обработку других поддеревьев, пока полностью не обработает текущее поддерево.

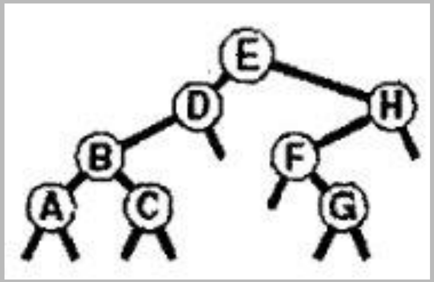
Для прохода по слоям в прямом или обратном порядке требуется другой алгоритм.

Обход дерева в ширину

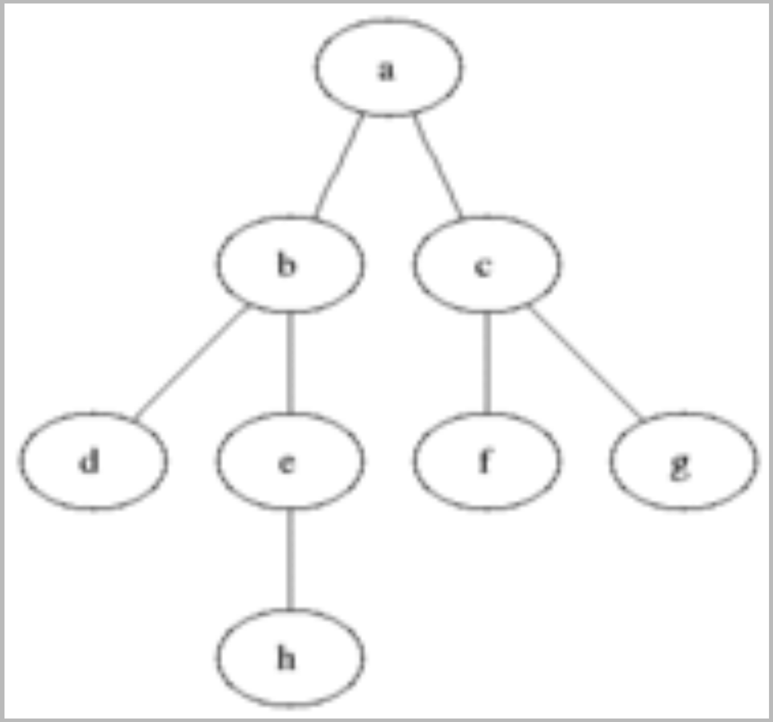
<u>Определение 12.</u> Обход двоичного дерева в ширину (BFS) — обход вершин дерева по уровням (слоям), начиная от корня.
BFS – Breadth First Search.

Используется очередь, в которой хранятся вершины, требующие просмотра.
За одну итерацию алгоритма:
* если очередь не пуста, извлекается вершина из очереди,
* посещается (обрабатывается) извлеченная вершина,
* в очередь помещаются все дочерние.

Порядок обработки узлов дерева на рис.:
E, D, H, B, F, A, C, G.



Обход дерева в ширину



Обход дерева в ширину

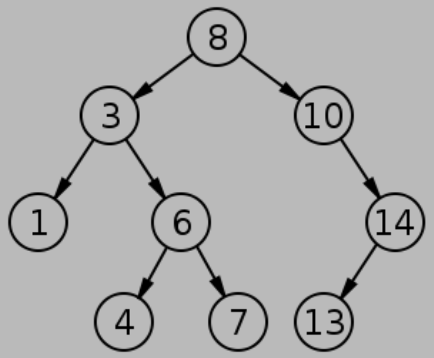
```
// Обход в ширину.
void TraverseBFS( CBinaryNode* root )
{
    queue<CBinaryNode*> q;
    q.put( root );
    while( !q.empty() ) {
        CBinaryNode* node = q.pop();
        visit( node );
        if( node->Left != NULL )
            q.push( node->Left );
        if( node->Right != NULL )
            q.push( node->Right );
    }
};
```

Двоичные деревья поиска

<u>Определение 13.</u> Двоичное дерево поиска (binary search tree, BST) – это двоичное дерево, с каждым узлом которого связан ключ, и выполняется следующее дополнительное условие:

- Ключ в любом узле X больше или равен ключам во всех узлах левого поддерева X и меньше или равен ключам во всех узлах правого поддерева X.

<u>Пример:</u>



Двоичные деревья поиска

Операции с двоичным деревом поиска:

1. Поиск по ключу.
2. Поиск минимального, максимального ключей.
3. Вставка.
4. Удаление.
5. Обход дерева в порядке возрастания ключей.

Двоичные деревья поиска

Поиск по ключу.

Дано: указатель на корень дерева X и ключ K.

Задача: проверить, есть ли узел с ключом K в дереве, и если да, то вернуть указатель на этот узел.

Алгоритм: Если дерево пусто, сообщить, что узел не найден, и остановиться.

Иначе сравнить K со значением ключа корневого узла X.

- Если K == X, выдать ссылку на этот узел и остановиться.
- Если K > X, рекурсивно искать ключ K в правом поддереве X.
- Если K < X, рекурсивно искать ключ K в левом поддереве X.

Время работы: O(h), где h – глубина дерева.

Двоичные деревья поиска

```
// Поиск. Возвращает узел с заданным ключом. NULL, если
узла
// с таким ключом нет.
CNode* Find( CNode* node, int value )
{
    if( node == NULL )
        return NULL;
    if( node->Data == value )
        return node;
    if( node->Data > value )
        return Find( node->Left, value );
    else
        return Find( node->Right, value );
};
```

Двоичные деревья поиска

Поиск минимального ключа.

<u>Дано</u>: указатель на корень непустого дерева X.

<u>Задача</u>: найти узел с минимальным значением ключа.

<u>Алгоритм</u>: Переходить в левый дочерний узел, пока такой существует.

Время работы: $O(h)$, где h – глубина дерева.

Двоичные деревья поиска

```
// Поиск узла с минимальным ключом.
CNode* FindMinimum( CNode* node )
{
    assert( node != NULL );
    while( node->Left != NULL )
        node = node->Left;
    return node;
};
```

Двоичные деревья поиска

Добавление узла.

Дано: указатель на корень дерева X и ключ K.

Задача: вставить узел с ключом K в дерево (возможно появление дубликатов).

Алгоритм: Если дерево пусто, заменить его на дерево с одним корневым узлом и остановиться.

Иначе сравнить K с ключом корневого узла X.

- Если $K < X$, рекурсивно добавить K в левое поддерево X.
- Иначе рекурсивно добавить K в правое поддерево X.

Время работы: $O(h)$, где h – глубина дерева.

Двоичные деревья поиска

```
// Вставка. Не указываем parent.
void Insert( CNode*& node, int value )
{
    if( node == NULL ) {
        node = new CNode( value );
        return;
    }
    if( node->Data > value )
        Insert( node->Left, value );
    else
        Insert( node->Right, value );
};
```

Двоичные деревья поиска

Удаление узла.

<u>Дано:</u> указатель на корень дерева X и ключ K.

<u>Задача:</u> удалить из дерева узел с ключом K (если такой есть).

<u>Алгоритм:</u> Если дерево пусто, остановиться.

Иначе сравнить K с ключом корневого узла X.

- Если K < X, рекурсивно удалить K из левого поддерева T.
- Если K > X, рекурсивно удалить K из левого поддерева T.
- Если K == X, то необходимо рассмотреть три случая:
  1. Обоих дочерних нет. Удаляем узел X, обнуляем ссылку.
  2. Одного дочернего нет. Переносим дочерний узел в X, удаляем узел.
  3. Оба дочерних узла есть.

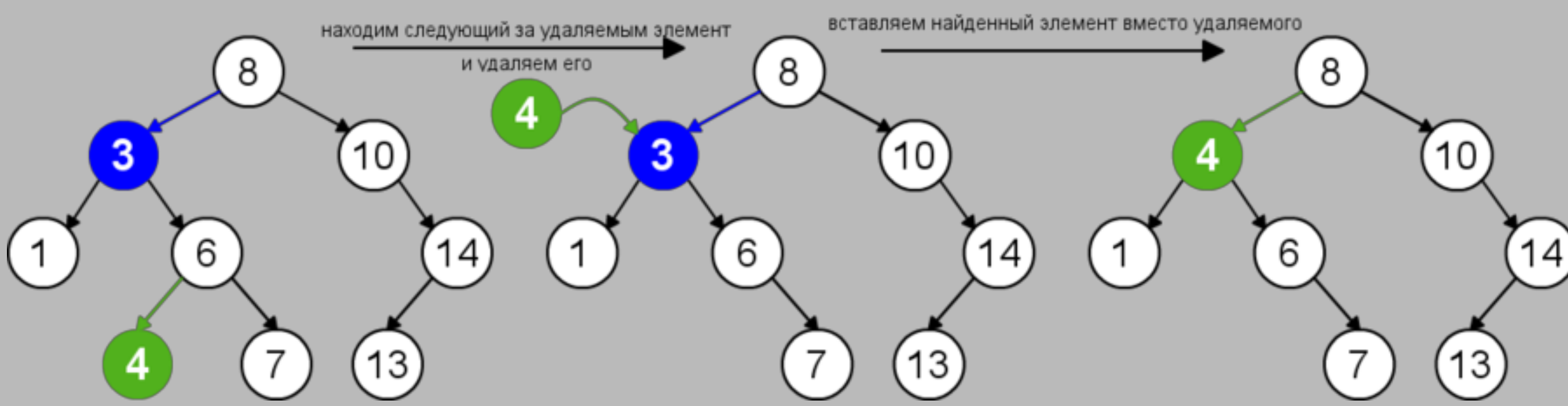
Двоичные деревья поиска

Удаление узла. Случай 3. Есть оба дочерних узла. Заменяем ключ удаляемого узла на ключ минимального узла из правого поддерева, удаляя последний.

Пусть удаляемый узел – X, а Y – его правый дочерний.

- Если у узла Y отсутствует левое поддерево, то копируем из Y в X ключ и указатель на правый узел. Удаляем Y.
- Иначе найдем минимальный узел Z в поддереве Y. Копируем ключ из Z, удаляем Z. При удалении Z копируем указатель на левый дочерний узел родителя Z на возможный правый дочерний узел Z.

Время работы удаления: O(h), где h – глубина дерева.



Двоичные деревья поиска

```
// Удаление. Возвращает false, если нет узла с заданным ключом.
bool Delete( CNode*& node, int value )
{
    if( node == 0 )
        return false;
    if( node->Data == value ) { // Нашли, удаляем.
        DeleteNode( node );
        return true;
    }
    return Delete( node->Data > value ?
        node->Left : node->Right, value );
};
```

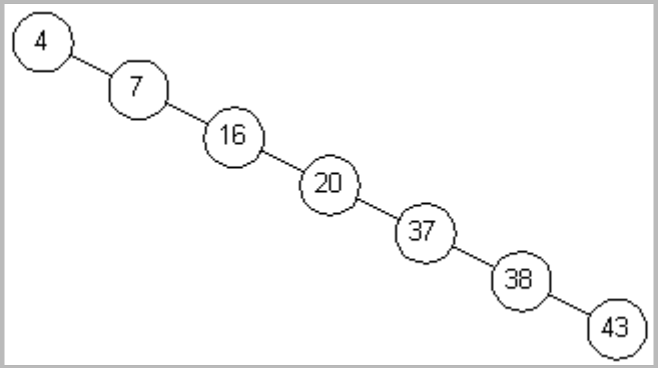
Двоичные деревья поиска

```
// Удаление узла.
void DeleteNode( CNode*& node )
{
    if( node->Left == 0 ) { // Если нет левого поддерева.
        CNode* right = node->Right; // Подставляем правое, может быть 0.
        delete node;
        node = right;
    } else if( node->Right == 0 ) { // Если нет правого поддерева.
        CNode* left = node->Left; // Подставляем левое.
        delete node;
        node = left;
    } else { // Есть оба поддерева.
        // Ищем минимальный элемент в правом поддереве и его родителя.
        CNode* minParent = node;
        CNode* min = node->Right;
        while( min->Left != 0 ) {
            minParent = min;
            min = min->Left;
        }
        // Переносим значение.
        node->Data = min->Data;
        // Удаляем min, подставляя на его место min->Right.
        (minParent->Left == min ? minParent->Left : minParent->Right)
            = min->Right;
        delete min;
    }
}
```

Балансировка

Все перечисленные операции с деревом поиска выполняются за $O(h)$, где h – глубина дерева.

Глубина дерева может достигать n. Последовательное добавление возрастающих элементов вырождает дерево в цепочку:



Необходима балансировка.

Балансировка

Самобалансирующиеся деревья.

Случайная балансировка:

- Декартовы деревья.

Гарантированная балансировка:

- АВЛ-деревья,
- Красно-черные деревья.

«Амортизированная» балансировка:

- Сплэй-деревья.

Декартовы деревья

Декартово дерево — это структура данных, объединяющая в себе двоичное дерево поиска и двоичную кучу.

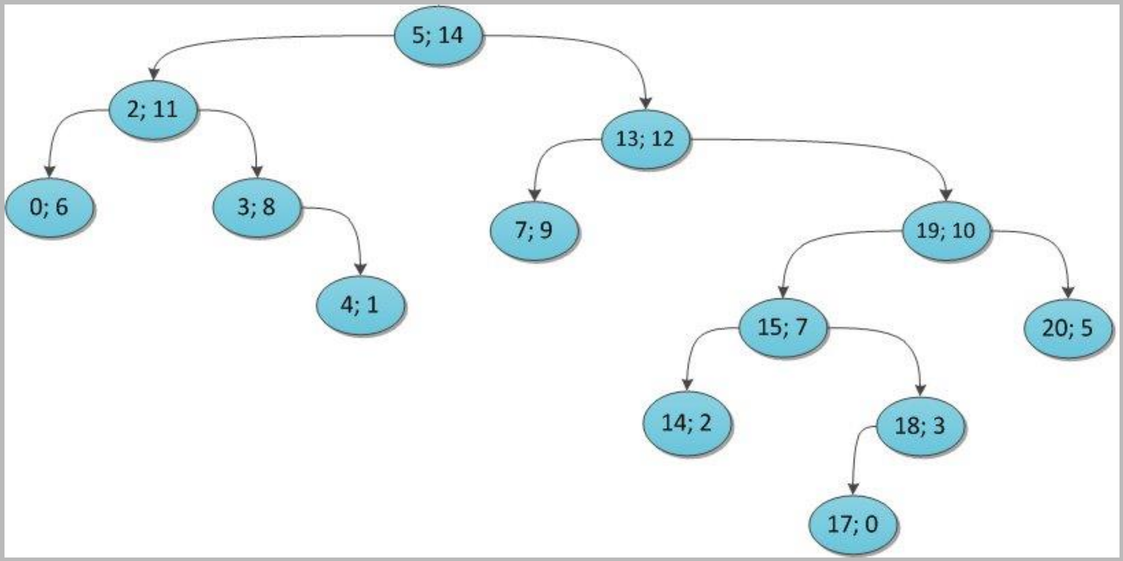
Определение 1. **Декартово дерево** — двоичное дерево, в узлах которого хранится пары (x, y), где x – это ключ, а y – это приоритет. Все x и все y являются различными. Если некоторый элемент дерева содержит (x_0, y_0), то у всех элементов в левом поддереве $x < x_0$, у всех элементов в правом поддереве $x > x_0$, а также и в левом, и в правом поддереве $y < y_0$.

Таким образом, декартово дерево является двоичным деревом поиска по x и кучей по y.

Декартовы деревья

Другие названия:
* treap (tree + heap),
* дуча (дерево + куча),
* дерамида (дерево + пирамида),
* курево (куча + дерево).

Изобретатели — Сидель и Арагон (1989г)



Декартовы деревья

Теорема 1. В декартовом дереве из n узлов, приоритеты которого являются случайными величинами с равномерным распределением, средняя глубина дерева $O(\log n)$.

Без доказательства.

Декартовы деревья

Основные операции:

- Разрезание – Split,
- Слияние – Merge.

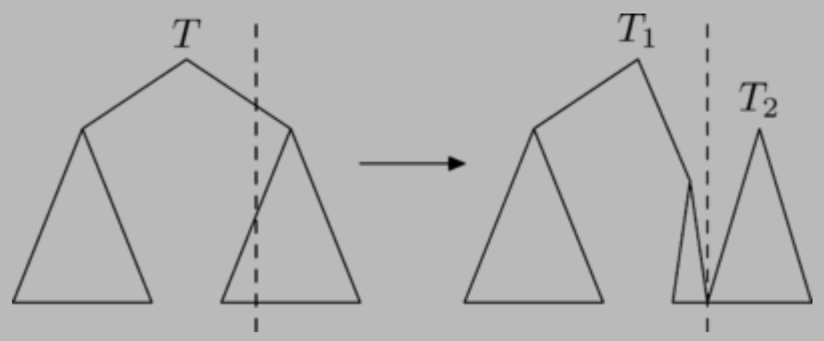
На основе этих двух операций реализуются операции:

- Вставка
- Удаление.

Декартовы деревья

Разрезание — Split

Операция **«разрезать»** позволяет разрезать декартово дерево T по ключу K и получить два других декартовых дерева: T_1 и T_2, причем в T_1 находятся все ключи дерева T, не большие K, а в T_2 – большие K.



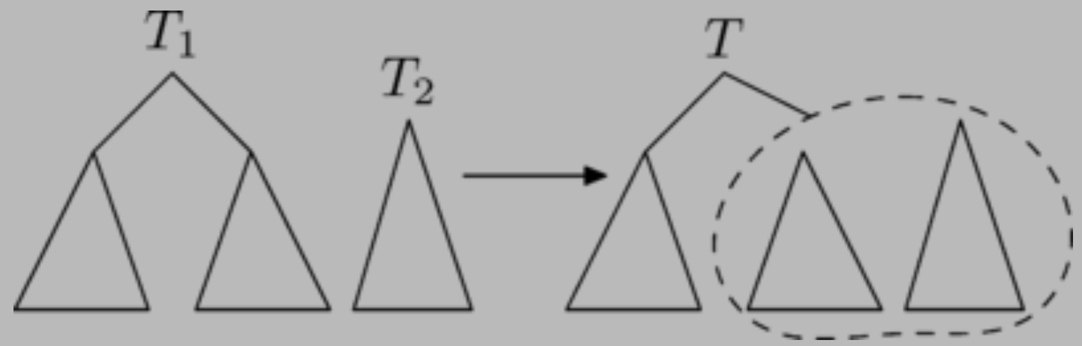
Декартовы деревья

```
// Разрезание декартового дерева по ключу.
void Split( CTreapNode* currentNode, int key, CTreapNode*& left,
    CTreapNode*& right )
{
    if( currentNode == 0 ) {
        left = 0;
        right = 0;
    } else if( currentNode->Key <= key ) {
        Split( currentNode->Right, key, currentNode->Right, right );
        left = currentNode;
    } else {
        Split( currentNode->Left, key, left, currentNode->Left );
        right = currentNode;
    }
}
```

Декартовы деревья

Слияние — Merge

Операция **«слить»** позволяет слить два декартовых дерева в одно. Причем, все ключи в первом (*левом*) дереве должны быть меньше, чем ключи во втором (*правом*). В результате получается дерево, в котором есть все ключи из первого и второго деревьев.



Декартовы деревья

```
// Слияние двух декартовых деревьев.
CTreapNode* Merge( CTreapNode* left, CTreapNode* right )
{
    if( left == 0 || right == 0 ) {
        return left == 0 ? right : left;
    }
    if( left->Priority > right->Priority ) {
        left->Right = Merge( left->Right, right );
        return left;
    }
    right->Left = Merge( left, right->Left );
    return right;
}
```

Декартовы деревья

Вставка

Добавляется элемент (x, y), где x – ключ, а y – приоритет.

Элемент (x, y) – это декартово дерево из одного элемента. Для того чтобы его добавить в наше декартово дерево T, очевидно, нужно их слить. Но T может содержать ключи как меньше, так и больше ключа x, поэтому сначала нужно разрезать T по ключу x.

<u>Реализация №1.</u>

1. Разобьём наше дерево по ключу x, который мы хотим добавить, на поддеревья T_1 и T_2.
2. Сливаем первое дерево T_1 с новым элементом.
3. Сливаем получившиеся дерево со вторым T_2.

Декартовы деревья

Вставка

Реализация №2.

1. Сначала спускаемся по дереву (как в обычном бинарном дереве поиска по x), но останавливаемся на первом элементе, в котором значение приоритета оказалось меньше y.
2. Теперь разрезаем поддерево найденного элемента на T_1 и T_2.
3. Полученные T_1 и T_2 записываем в качестве левого и правого сына добавляемого элемента.
4. Полученное дерево ставим на место элемента, найденного в первом пункте.

В первой реализации два раза используется Merge, а во второй реализации слияние вообще не используется.

Декартовы деревья

Удаление.

Удаляется элемент с ключом x.

<u>Реализация №1.</u>

1. Разобьём дерево по ключу x, который мы хотим удалить, на T_1 и T_2.
2. Теперь отделяем от первого дерева T_1 элемент x, разбивая по ключу $x - \varepsilon$.
3. Сливаем измененное первое дерево T_1 со вторым T_2.

Декартовы деревья

Удаление.

Реализация №2.

1. Спускаемся по дереву (как в обычном двоичном дереве поиска по x), ища удаляемый элемент.
2. Найдя элемент, вызываем слияние его левого и правого сыновей.
3. Результат процедуры ставим на место удаляемого элемента.

В первой реализации два раза используется Split, а во второй реализации разрезание вообще не используется.

Декартовы деревья

Расход памяти и время работы.

	В любом случае	В среднем случае	В худшем случае
Расход памяти	$O(n)$	$O(n)$	$O(n)$
Поиск	$O(h)$	$O(\log n)$	$O(n)$
Вставка	$O(h)$	$O(\log n)$	$O(n)$
Удаление	$O(h)$	$O(\log n)$	$O(n)$

АВЛ-дерево

<u>Определение.</u> АВЛ-дерево — сбалансированное двоичное дерево поиска. Для каждой его вершины высоты её двух поддеревьев различаются не более чем на 1.

Изобретено Адельсон-Вельским Г.М. и Ландисом Е.М. в 1962г.

АВЛ-дерево

Теорема. Высота АВЛ-дерева $h = O(\log n)$.

Идея доказательства. В АВЛ-дереве высоты h не меньше F_h узлов, где F_h – число Фибоначчи.
Из формулы Бине следует, что

$$n \geq F_h = \frac{\phi^h - (-\phi)^{-h}}{\phi - (-\phi)^{-1}} \geq C\phi^h,$$

где $\phi = \left(1 + \sqrt{5}\right)/2$ – золотое сечение.

АВЛ-дерево

Специальные балансирующие операции, восстанавливающие основное свойство «высоты двух поддеревьев различаются не более чем на 1» – вращения.

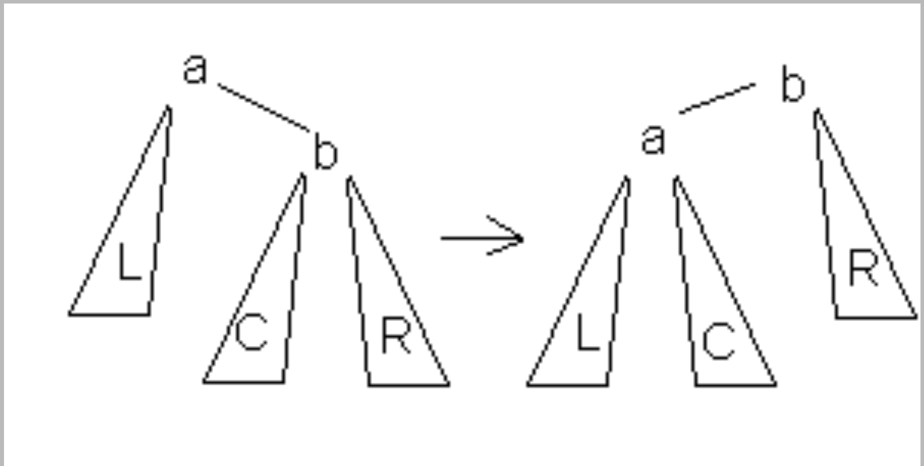
- Малое левое вращение,
- Малое правое вращение,
- Большое левое вращение,
- Большое правое вращение.

АВЛ-дерево

- **Малое левое вращение**

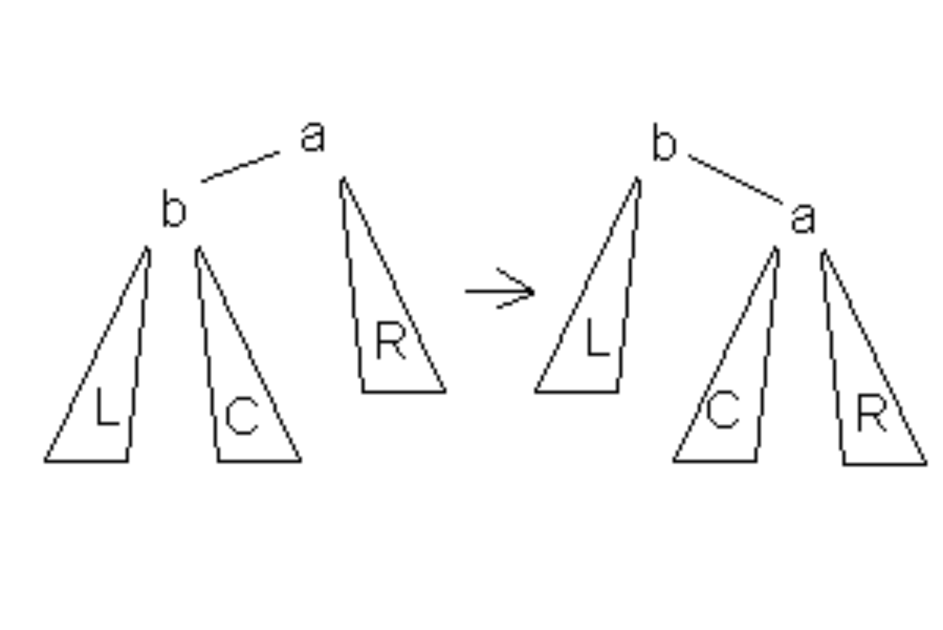
Используется, когда:
высота(R) = высота(L) + 2 и высота(C) ≤ высота(R).

После операции:
высота дерева останется прежней, если высота(C) = высота(R),
высота дерева уменьшится на 1, если высота(C) < высота(R).



АВЛ-дерево

- Малое правое вращение

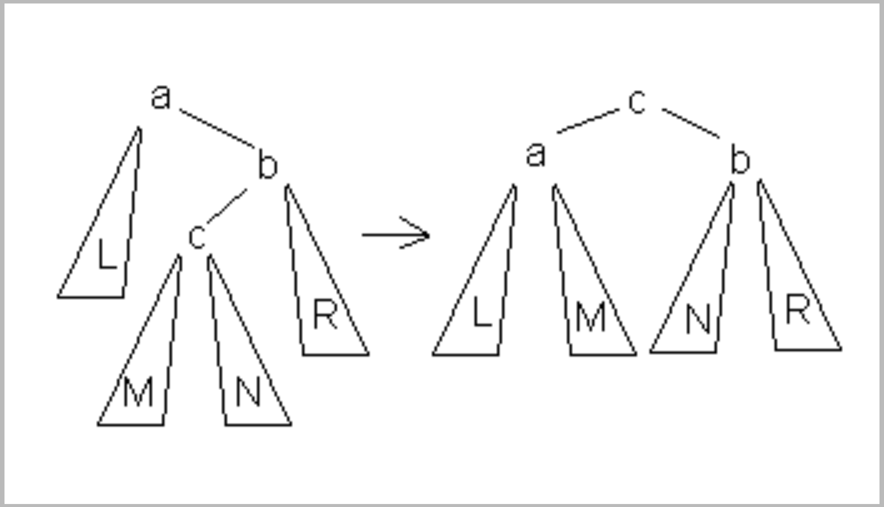


АВЛ-дерево

- **Большое левое вращение**

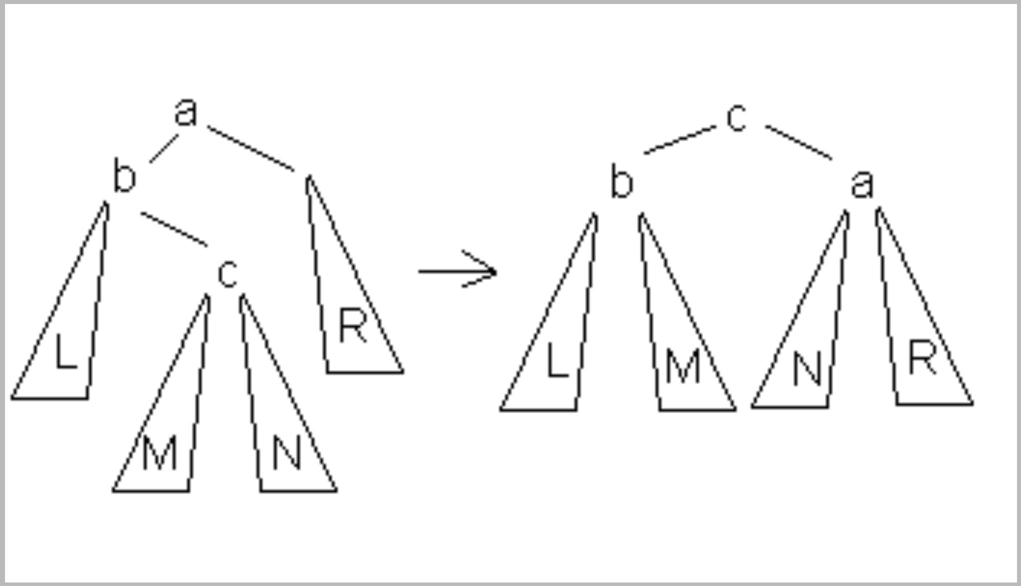
Используется, когда:
высота(R) = высота(L) + 1 и высота(C) = высота(L) + 2.

После операции:
высота дерева уменьшается на 1.



АВЛ-дерево

- **Большое правое вращение**



АВЛ-дерево

Вставка элемента

1. Проходим по пути поиска, пока не убедимся, что ключа в дереве нет.
2. Включаем новую вершину как в стандартной операции вставки в дерево поиска.
3. "Отступаем" назад от добавленной вершины к корню. Проверяем в каждой вершине сбалансированность. Если разность высот поддеревьев равна 2 – выполняем нужное вращение.

Время работы = $O(\log n)$.

АВЛ-дерево

Удаление элемента

1. Ищем вершину D, которую требуется удалить.
2. Проверяем, сколько поддеревьев в D:
   - Если D – лист или D имеет одно поддерево, то удаляем D.
   - Если D имеет два поддерева, то ищем вершину M, следующую по значению после D. Как в стандартном алгоритме удаления из дерева поиска. Переносим значение из M в D. Удаляем M.
3. "Отступаем" назад от удаленной вершины к корню. Проверяем в каждой вершине сбалансированность. Если разность высот поддеревьев равна 2 – выполняем нужное вращение.

Время работы = $O(\log n)$.

АВЛ-деревья

Расход памяти и время работы.

	В среднем случае	В худшем случае
Расход памяти	$O(n)$	$O(n)$
Поиск	$O(\log n)$	$O(\log n)$
Вставка	$O(\log n)$	$O(\log n)$
Удаление	$O(\log n)$	$O(\log n)$

АТД «Ассоциативный массив»

<u>Определение 4.</u> Ассоциативный массив — абстрактный тип данных, позволяющий хранить пары вида «(ключ, значение)» и поддерживающий операции добавления пары, а также поиска и удаления пары по ключу:

- INSERT(ключ, значение).
- FIND(ключ). Возвращает значение, если есть пара с заданным ключом.
- REMOVE(ключ).

Предполагается, что ассоциативный массив не может хранить две пары с одинаковыми ключами.

АТД «Ассоциативный массив»

Расширение ассоциативного массива.

Обязательные три операции часто дополняются другими. Наиболее популярные расширения включают следующие операции:

- CLEAR – удалить все записи.
- EACH – «пробежаться» по всем хранимым парам
- MIN – найти пару с минимальным значением ключа
- MAX – найти пару с максимальным значением ключа

В последних двух случаях необходимо, чтобы на ключах была определена операция сравнения.

АТД «Ассоциативный массив»

Реализации ассоциативного массива.

- Массив пар, упорядоченный по ключу. Поиск – бинарный.

Время поиска $O(\log n)$. Время вставки и удаления $O(n)$.

- Сбалансированное дерево поиска.

Время работы операций поиска, вставки и удаления – $O(\log n)$.

std::map реализован на основе красно-черного дерева.

- Хеш-таблицы.

Все операции в среднем – $O(1)$, в худшем – $O(n)$.

МГТУ им. Н.Э. Баумана
ТЕХНОПАРК
Mail.Ru Group

Мацкевич С.Е.

Спасибо за внимание!