# Реализация алгоритов Ху—Таккера и Гарсия—Уочса для построения оптимальных алфавитных деревьев

## Теоретическая часть

Алфавитное дерево — это такое бинарное дерево, при просмотре листьев которого слева направо, соответствующие буквы должны появляться в алфавитном порядке.

### 1. Алгоритм Ху—Таккера

Пусть имеется порядок листьев и их веса . Эти последовательности назовём соответственно последовательностью узлов и весов.



Если мы комбинируем и , то их отец обозначается через , его вес — через . Если комбинируется и , их отец обозначается через . В этих



Алгоритм Ху—Таккера сначала строит дерево, не являющееся алфавитным, а затем преобразует его в оптимальное алфавитное дерево.

**Понятия:**

Два узла в последовательности узлов называются *совместимой парой*, если они соседние или все узлы между ними круглые (т. е. имеют потомков).

Когда комбинируется пара узлов с весами и , вес их отца называется *весом этой пары*. Пара с минимальным весом называется *минимальной парой*.



Алгоритм Ху—Таккера строит алфавитное дерево, минимизирующее , выполняя следующие шаги.



Комбинирование. По данной последовательности из n узлов с весами строим последовательность из n – 1 узла, комбинируя локально минимальную совместимую пару (пара, вес которой меньше, чем вес соседних с ней совместимых пар), заменяя левого сына его отцом и удаляя правого сына из последовательности. Процедура слияния весов продолжается до тех пор, пока не останется один вес.



Определение уровней. Находим номер уровня каждого листа относительно корня (узел с наибольшим значением располагается внизу дерева).



Перестройка. После того, как номера уровней всех листьев определены, применим ним стековый алгоритм. Он заключается в следующих шагах.



Шаг 0. Стек пуст, находятся в очереди.



Шаг 1. Если в стеке меньше двух элементов, перейти к шагу 2. В противном случае проверить, равны ли значения двух верхних элементов стека. Если они различны, перейти к шагу 2, а если равны — к шагу 3.

Шаг 2. Удалить из очереди первый элемент и поместить его на вершину стека. Перейти к шагу 1.

Шаг 3. Пусть – верхний элемент стека, а — следующий элемент. Заменить и на . Если , остановиться, иначе перейти на шаг 1. (Это означает, что комбинируются узлы и , а их отец становится узлом уровня .)



### 2. Алгоритм Гарсиа—Уочса

В фазе комбинирования алгоритма Ху—Таккера мы последовательно комбинируем л. м. с. п., при этом рассматриваемые пары могут быть разделены несколькими узлами-отцами. Алгоритм Гарсиа—Уочса устраняет различия между листьями и узлами и располагает узлы в последовательность так, что л. м. с. п. всегда является соседней парой. В последовательности листьев соседняя пара есть л. м. с. п. тогда и только тогда, когда

и



Пусть — последовательность весов. Опишем для неё алгоритм Гарсиа—Уочса.



Найти самую левую минимальную соседнюю пару, .



Скомбинировать и в один узел с весом .



Передвинуть влево, пропуская все узлы, вес которых меньше или равен . Получить новую рабочую последовательность из n – 1 узла



Где .



Этот процесс повторяется, пока в последовательности узлов не останется один узел. Тем самым будет построено дерево из первого шага алгоритма Ху—Таккера. Остальная часть такая же, как в алгоритме Ху—Таккера.