* Что такое словарь, ассоциативный массив?

Словарь (dictionary) — это структура данных для хранения пар вида «ключ» — «значение» (key — value). Альтернативное название — ассоциативный массив (associative array, map). В словаре может быть только одна пара с заданным ключом.

* Что такое бинарное дерево поиска? Проведите анализ сложности основных операций.

Бинарное дерево (binary tree) — это дерево (структура данных), в котором каждый узел (node) имеет не более двух дочерних узлов (child nodes).

Бинарное дерево поиска (двоичное дерево поиска, binary search tree, BST) — это бинарное дерево, в котором: ➔ Каждый узел x имеет не более двух дочерних узлов и содержит пару «ключ» — «значение»

➔ Ключи всех узлов левого поддерева x меньше значения его ключа

➔ Ключи всех узлов правого поддерева x больше значения его ключа

В основном операции для бинарных деревьев поиска имеют сложность O(h), где h — высота дерева.

* Что такое хеш-таблица? Проведите анализ сложности основных операций

Хеш-таблица (hash table) — это структура данных для хранения пар «ключ» — «значение».

* Доступ к элементам осуществляется по ключу (key).
* Ключи могут быть строками, числами, указателями и т.д.
* Ограничение — ключи (индексы) могут быть только целыми неотрицательными числами.

Хеш-таблицы позволяют в среднем за время O(1) выполнять добавление, поиск и удаление узлов.

* Что такое хеш-функция? Какая хеш-функция является «хорошей»?

Хеш-функция (hash function) — это функция, преобразующая значение ключа (например, строки, числа, файла) в целое число. Значение, возвращаемое хеш-функцией, называется хеш-кодом (hash code), контрольной суммой (hash sum) или просто хешем (hash).

Время вычисления хеш-функции зависит от длины ключа и не зависит от количества элементов в массиве.

«Хорошая» хеш-функция должна удовлетворять двум свойствам:

* быстрое вычисление (за линейное от размера объекта время);
* минимальное количество «коллизий» (совпадение значений хеш-функции для двух разных ключей).
* Методы разрешения коллизий в хеш-таблицах.

Широкое распространение получили два основных метода разрешения коллизий (collision resolution): метод цепочек (chaining) и открытая адресация (open addressing).

В *методе цепочек* каждая ячейка хеш-таблицы содержит указатель на голову связного списка, в который помещаются все ключи, имеющие одинаковый хеш-код ℎ. В каждом узле списка содержатся ключ и некоторое значение.

Основная идея *открытой адресации* – хранение ключей и ассоциированных с ними значений непосредственно в ячейках таблицы. Такой подход позволяет более экономно использовать память.

Коллизия

Коллизия в хеш-таблицах — это конфликт, возникающий, когда хеш-функция генерирует один индекс для нескольких ключей. В этом случае неизвестно, какое значение нужно сохранить в этом индексе.

Сам подсчёт коллизий происходит в функции добавления, если у нас встречается такой случай при котором hashtab[index] != NULL, то f = 1 и мы прибавляем к colission\_count единицу





где m это длинна ключа, n кол-во элементов, а h это высота таблицы