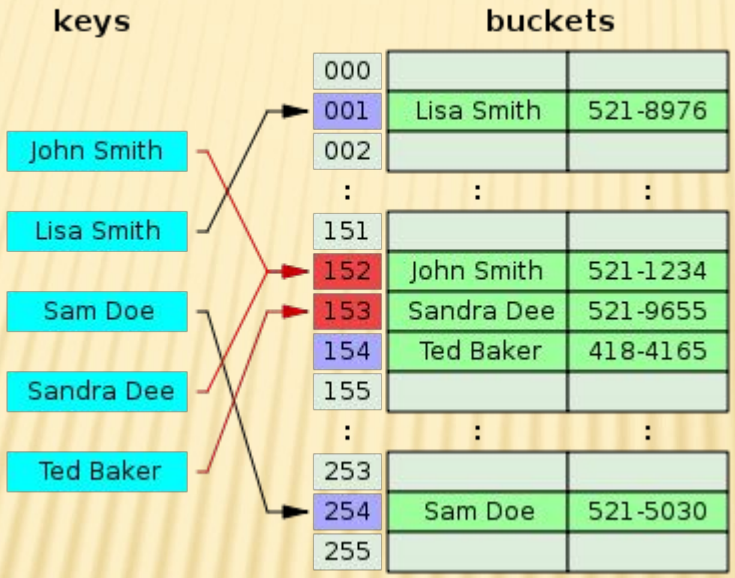
# ХЕШИРОВАНИЕ ДАННЫХ

В настоящее время широко используется распространенный метод обеспечения быстрого доступа к большим объемам информации – хеширование.

Хеширование (или хэширование, англ. hashing)– это преобразование входного массива данных определенного типа и произвольной длины в выходную битовую строку фиксированной длины. Такие преобразования также называются хеш-функциями или функциями свертки, а их результаты называют хешем, хеш-кодом, хеш-таблицей или дайджестом сообщения (англ. message digest).

Хеш-таблица–это структура данных, реализующая интерфейс ассоциативного массива, то есть она позволяет хранить пары вида "ключ-значение" и выполнять три операции: операцию добавления новой пары, операцию поиска и операцию удаления пары по ключу. Хеш-таблица является массивом, формируемым в определенном порядке хеш-функцией.




Пример хеш-таблицы с открытой адресацией

Принято считать, что хорошей, с точки зрения практического применения, является  такая хеш-функция, которая удовлетворяет следующим условиям:

- функция должна быть простой с вычислительной точки зрения;
- функция должна распределять ключи в хеш-таблице наиболее равномерно;
- функция не должна отображать какую-либо связь между значениями ключей в связь между значениями адресов;
- функция должна минимизировать число коллизий – то есть ситуаций, когда разным ключам соответствует одно значение хеш-функции (ключи в этом случае называются синонимами).
- При этом первое свойство хорошей хеш-функции зависит от характеристик компьютера, а второе – от значений данных.

Хеш-таблицы должны соответствовать следующим свойствам:

- Выполнение операции в хеш-таблице начинается с вычисления хеш-функции от ключа. Получающееся хеш-значение является индексом в исходном
- массиве.

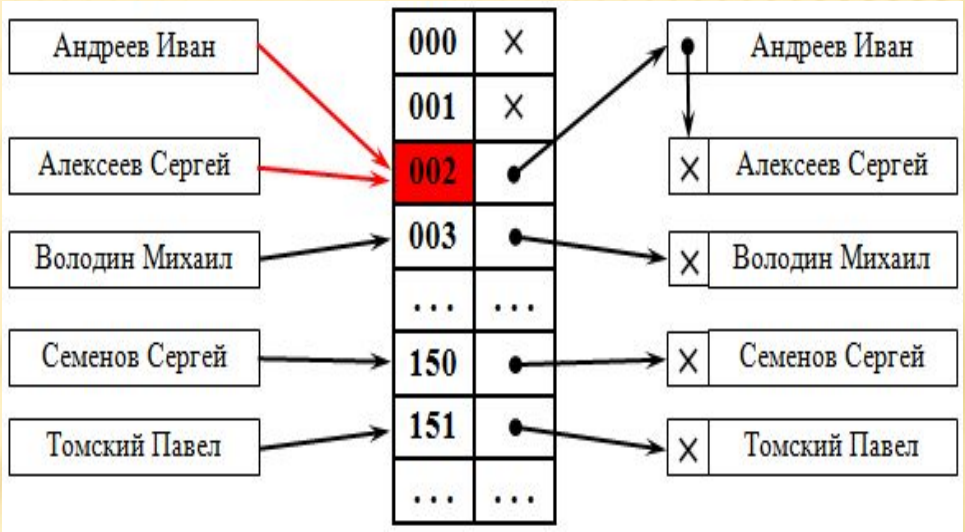
Количество хранимых элементов массива, деленное на число возможных значений хеш-функции, называется коэффициентом заполнения хеш-таблицы (load factor) и является важным параметром, от которого зависит среднее время
- выполнения операций.

Операции поиска, вставки и удаления должны выполняться в среднем за время O(1). Однако при такой оценке не учитываются возможные аппаратные затраты на перестройку индекса хеш-таблицы, связанную с увеличением значения размера массива и добавлением в хеш-таблицу новой пары.

Механизм разрешения коллизий является важной составляющей любой хеш-таблицы. Коллизии осложняют использование хеш-таблиц, так как нарушают однозначность соответствия между хеш-кодами и данными.

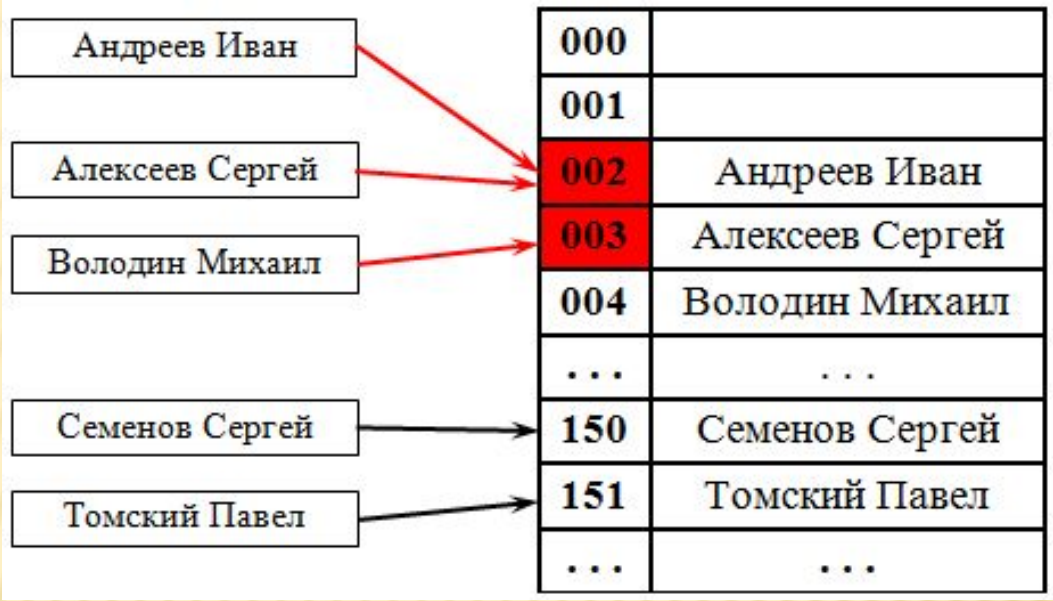
Тем не менее, существуют способы преодоления возникающих сложностей:

- метод цепочек (внешнее или открытое хеширование);




Каждая ячейка массива является указателем на связный список (цепочку) пар ключ-значение, соответствующих одному и тому же хеш-значению ключа. Коллизии просто приводят к тому, что появляются цепочки длиной более одного элемента.

- метод открытой адресации (закрытое хеширование);




Если ячейка с вычисленным индексом занята, то можно просто просматривать следующие записи таблицы по порядку до тех пор, пока не будет найден ключ K или пустая позиция в таблице.

Хеширование имеет широкое практическое применение в теории баз данных, кодировании, банковском деле, криптографии и других областях.

**Работу выполнил студент группы СИС-11:**
**Филиппов Андрей Юрьевич**