

Лабораторная работа №5.

Хэширование.

1 Теоретические сведения

Хэш-таблицы представляют очень практичный способ реализации словарей. В них используется то обстоятельство, что поиск элемента по индексу выполняется за постоянное время. Хэш-функция соотносит набор ключей с целочисленными значениями. Мы будем использовать значение хэш-функции в качестве индекса массива и записывать элементы в этой позиции. Сначала хэш-функция соотносит каждый ключ с большим целым числом. Пусть значение a представляет размер алфавита, используемого для создания строки S . Пусть $char(c)$ будет функцией, которая однозначно отображает каждый символ алфавита в целое число от 0 до $a - 1$. Функция $\sum_{i=0}^{|S|-1} a^{|S|-(i+1)} \times char(S_i)$ однозначно отображает целую строку в (большое) целое число, рассматривая символы строки как цифры системы счисления с основанием a .

В результате получаются уникальные идентификационные числа, но они будут настолько большими, что очень быстро превысят количество ячеек в нашей хэш-таблице (обозначаемое m). Это значение необходимо уменьшить до целого числа в диапазоне от 0 до $m - 1$, для чего выполняется операция получения остатка от деления.

Независимо от того насколько хороша наша хэш-функция, время от времени она будет отображать два разных ключа в одно хэш-значение, и нужно быть готовым к подобной ситуации. Самым легким способом разрешения таких коллизий является применение цепочек. Для этого хэш-таблица реализуется в виде массива из m связанных списков. Список с порядковым номером содержит все элементы, хэшированные в одно и то же значение i .

2 Задание

Реализуйте класс “телефонный справочник”. В справочнике должны храниться фамилии людей и их номера телефонов. Должна быть возможность получения номера телефона по фамилии человека. Для реализации класса используйте хэш-функцию.

Контрольные вопросы

1. Приведите пример применения словарей.
2. Как строится хэш-таблица?
3. Для чего, кроме словарей, может применяться хэш-таблица.