Хеш-функция

Билет № ?

11 декабря 2025 г.

1 Математическое ожидание

Будем говорить, что случайная величина ξ имеет равномерное распределение на конечном множестве $X \subset \mathbb{R}$, если

$$\mathbb{P}\{\xi = x\} = \frac{1}{|X|}, \quad \forall x \in X.$$

Математическое ожидание – это взвешенная сумма произведения значения случайной величины на вероятность выпадения случайной величины. Математическое ожидание обозначается $\mathbb{E}\xi$ и задается следующей формулой:

$$\mathbb{E}\xi = \sum_{\omega \in \Omega} \xi(\omega)\mathbb{P}\{\omega\}.$$

2 Хеш-функция

- **Определение. Хеш-функцией** будем называть отображение h: K -> {0, 1, ..., m - 1}, где m $\in \mathbb{N}$.

- **Определение. Коллизия.** Будем говорить, что ключи a != b образуют коллизию, если h(a) = h(b).

Пусть у нас есть хеш-функция h. Инициализируем массив T размера m. В ячейках будем хранить двухсвязные списки. Теперь, когда мы хотим вставить элемент, мы будем вставлять его в начало двухсвязного списка

T[h(k)], где k - ключ k(x). Таким образом можно разрешать коллизии, храня элементы с одинаковым хешлом в одном списке.

Замечание. Поиск элемента с ключом k сводится к поиску в двухсвязном списке T[h(k)].

2.1 Гипотеза о простом равномерном хешировании

Пусть хеш-таблица содержит n элементов.

Определение. Load-factor. α := n / m.

2.1.1 Гипотеза.

Пусть дана хеш-функция h: $\mathbb{P}$ -> {0,1,2,...m - 1} и случайная величина $k : \Omega \to \mathbb{P}$. Тогда h удовлетворяет гипотезе простого равномерного хеширования относительно k, если h распределяет ключи по ячейкам независимо и равномерно, т.е.:

- $\forall\ h_0 \in 0, 1, ...(m-1)$ и $\forall k \to \mathbb{P}\{\text{h(k)} = h_0\} = 1/\text{m}$

Фан факт Для любых случайных величин $X_1, X_2, \ldots, X_n$ (даже зависимых!) и любых констант $a_1, a_2, \ldots, a_n$ выполняется: $$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i]$$

Утверждение. Длина списка l, отвечающая хешу h_0 в среднем равна α.

Доказательство:

Занумеруем ключи. ТОгда длина списка равна:

$$\mathbb{E}\left[\sum_{p=1}^{n} \mathbb{I}[h(k_p) = h_0]\right] = \sum_{p=1}^{n} \mathbb{E}\left[\mathbb{I}[h(k_p) = h_0]\right] = \sum_{p=1}^{n} \mathbb{P}\left[h(k_p) = h_0\right] = \sum_{p=1}^{n} \frac{1}{m} = \frac{n}{m} = \alpha$$

Когда берем матожидание от I происходит перемножение:

*1 * $P(h(k_p) = h_0)$ + 0 * $P(h(k_p)! = h_0)$ = $P(h(k_p) = h_0)$*

Ценок.

Следствие. Если n = $O(m)$, то время работы поиска в среднем $\Theta(1)$

Доказательство:

$\Theta(\alpha) = \Theta(\frac{n}{m}) = \Theta(\frac{m}{m}) = \Theta(1)$

Ценок.