

МФТИ, ФПМИ  
Алгоритмы и структуры данных, 2-й семестр, весна 2022  
Семинар №4. Хеш-таблицы

1. Реализуйте структуру данных, которая поддерживает следующие операции:

- добавить пару (ключ, значение);
- удалить пару по ключу;
- по ключу вернуть значение;
- сообщить текущее количество пар;
- выбрать равновероятно случайную пару (ключ, значение);
- вывести все хранящиеся на текущий момент пары в любом порядке за время  $O(n)$ , где  $n$  — их количество.

Все операции, кроме последней, должны работать за  $O^*(1)$  амортизированно.

2. Реализуйте структуру LRU Cache, которая имеет фиксированный размер и поддерживает следующие операции:

- Добавить пару (ключ, значение). Если количество пар превысило ограничение, нужно удалить ту пару из текущего набора пар, к которой обращались раньше всего, то есть *least recently used*.
- По ключу вернуть значение.

3. Вам даны два массива одинаковой длины, состоящие из  $n$  чисел. Проверьте, что они задают одинаковые множества, за  $O(n)$  времени в среднем и  $O(n)$  памяти. Решите задачу в предположении:

- а) все числа различны;
- б) числа могут повторяться.

4. Найдите пересечение двух неотсортированных массивов за  $O(n)$  времени в среднем и  $O(n)$  памяти.

5. Дан массив, состоящий из  $n$  чисел. Найдите в нём самую длинную подпоследовательность последовательных целых чисел. Например, для массива  $[1, 5, 2, 7, 3]$  ответ —  $[1, 2, 3]$ . Асимптотика:  $O(n)$  в среднем.

6. Какая связь между понятиями *2-независимое семейство* и *универсальное семейство*?

7. Назовём семейство хеш-функций равномерно распределённым, если при равномерном случайном выборе хеш-функции случайная величина, равная значению хеш-функции на фиксированном элементе  $x$ , распределена на пространстве значений равномерно. Далее, 2-независимым семейством хеш-функций называется семейство, в котором описанные выше случайные величины для любых двух различных фиксированных  $x$  и  $y$  независимы. Приведите пример семейства хеш-функций, которое:

- а) равномерно распределено, но не является 2-независимым;
- б) является 2-независимым, но не является равномерно распределённым.

8. Пусть  $\mathbb{F}$  — конечное поле размера  $f$ . Выберем  $\alpha$  и  $\beta$  равновероятно из  $\mathbb{F}$ . Определим  $h_{\alpha, \beta}(x) = \alpha x + \beta$  как функцию над  $\mathbb{F}$ .

- а) Докажите, что  $\{h_{\alpha, \beta}(u) \mid u \in \mathbb{F}\}$  — семейство равномерно распределённых попарно независимых случайных величин.
- б) Обобщите этот подход на многочлены степени  $k - 1$ . Докажите, что получившееся семейство является  $k$ -независимым.

9. Почему в фильтре Блума с лекции нельзя удалить ключ, просто занулив соответствующие биты? Придумайте, как изменить фильтр Блума, чтобы из него можно было удалять элементы. При этом придётся отказаться от экономии памяти.

1. Заведите вместе с хеш-таблицей вектор всех пар. Каждая ячейка хеш-таблицы будет указывать на некоторый элемент вектора.
2. Вновь используйте вектор.
3. Инициализируйте совершенное хеширование на элементах первого массива и прогоните через него второй массив.
4. Всё то же самое.
5. Вспомните, как мы искали наибольшую возрастающую подпоследовательность.
6. Из 2-независимости следует универсальность.
7.
  - а) Подойдёт, например, семейство функций, каждая из которых является константой.
  - б) Подойдёт, например, такое семейство: значения всех хеш-функций на всех элементах, кроме одного, фиксированы, а один принимает всевозможные значения в разных функциях. Получится семейство размера  $M$ .
8. Сведите задачу к решению системы линейных уравнений. Решение всегда существует для различных  $u$  и  $v$ . В общем случае получится матрица Вандермонда, определитель которой будет ненулевым, т.к. все  $u_i$  различны.
9. Вместо битов используйте счётчики.