

## Хеширование (06.07.2009, группа «П», Илья Разенштейн)

1. Некоторые сведения из теории вероятностей: неравенство для объединения, условные вероятности, независимые события, независимые случайные величины, математическое ожидание, его свойства (линейность, неравенство Маркова).
2. Хеширование как эвристическое ускорение наивного поиска.
3. Основные реализации хеширования:
  - (a) открытая адресация,
  - (b) хеширование со списками,
  - (c) хеширование подстрок, замена суффиксных структур.
4. Фильтры Блюма. Их применение на практике.
5. Формализация понятия «случайная хеш-функция». Почему мы не можем позволить себе хранить случайную хеш-функцию. Универсальные семейства хеш-функций:
  - (a) универсальное семейство в слабом смысле,
  - (b) универсальное семейство в сильном смысле,
  - (c)  $k$ -универсальное семейство.

Формулировка результатов, связанных с существованием универсальных семейств.
6. Анализ времени работы хеш-таблицы со списками при использовании универсального семейства в слабом смысле.
7. Статический словарь. Как получить время доступа  $O\left(\frac{\log n}{\log \log n}\right)$ ,  $O(\log \log n)$  в худшем случае.
8. Хеширование без коллизий. Словарь FKS.
9. Применение хеширования в теории: сведение Валианта-Вазирани.

## Домашнее задание

1. Приведите пример вероятностного пространства  $\Omega$  и трех событий в нем  $A_1$ ,  $A_2$  и  $A_3$  таких, что они являются попарно независимыми, но не являются независимыми в совокупности.
2. Докажите, что подмножество независимых случайных величин является независимым.
3. Проанализируйте фильтр Блюма и расскажите, как вы бы его использовали (предполагается подбор параметров  $m$  и  $k$ ). Что можно сказать про удаление элементов?
4. Докажите, что семейство линейных отображений  $x \mapsto Ax + b$  является универсальным в сильном смысле.
5. Докажите, что математическое ожидание максимальной длины цепочки при использовании  $n$ -универсального семейства —  $O\left(\frac{\log n}{\log \log n}\right)$ .