

1. (Парадокс дней рождения) Пусть $|B| \geq |U|^2$, а функция $h : U \rightarrow B$ на элементах U принимает независимые значения, равномерно распределённые в B . Покажите, что $\mathbb{P}(\exists x \neq y : h(x) = h(y)) \leq 1/2$. Убедитесь, что чем больше $|B|$, тем меньше вероятность коллизии.
2. Китайская теорема об остатках гласит следующее: если m_1, \dots, m_k попарно взаимно просты, то для произвольного набора целых чисел r_1, \dots, r_k система сравнений $\{x \equiv r_i \pmod{m_i}\}_{i=1}^k$ имеет единственное решение среди чисел $\{0, 1, \dots, p-1\}$, где $p = m_1 \cdot \dots \cdot m_k$. Как можно расширить размер множества хешей до чисел порядка 10^{40} ? *Указание:* Используйте несколько хеш-функций.
3. Почему префикс-функцию нельзя искать с помощью бинарного поиска и проверки равенства подстрок хешами? Можно ли так считать z -функцию?
4. Алгоритм Кнута—Морриса—Пратта ищет все вхождения шаблона s в текст t . Для этого рассматривается префикс-функция строки $s\#t$. Покажите, зачем нужен разделитель $\#$. Покажите, как оптимизировать алгоритм для потребления $O(|s|)$ памяти.
5. Строгим периодом строки s называется такая строка t минимальной длины, что существует целое $k \geq 1$, т. ч. $t^k = s$. Нестрогим периодом называется такая t минимальной длины, что существует целое $m \geq 1$, т. ч. $s \sqsubset t^m$ (то есть s является префиксом t^m). Найдите строгий и нестрогий периоды s за $O(|s|)$.
6. В строке s за $O(|s|^2)$ найдите количество её различных подстрок, то есть размер множества $\{s_l \dots s_r \mid l \leq r\}$.
7. По z -функции строки постройте её префикс-функцию.
8. В данной строке s за $O(|s|)$ предподсчёта научитесь отвечать на запрос “равна ли подстрока $s_l \dots s_r$ префиксу соответствующей длины?” за $O(1)$.
9. В данной строке найдите все префиксы, являющиеся палиндромами.
10. Дан шаблон p и текст t . Найдите все вхождения p в t не более чем с одной ошибкой (то есть нужно найти подстроки t , которые равны p с точностью до замены одного символа).
- 11*. По данной префикс-функции строки найдите количество строк над алфавитом мощности m с той же префикс-функцией.