

Разбор конкурса № 15 "Хэши"

Игорь Демидов

19 декабря 2012

1 Подстроки

Первый шаг решения - бинарный поиск по ответу. Зафиксируем длину наибольшей общей подстроки, которую мы ищем. Пусть она равна L . Тогда для каждой из K строк посчитаем хэши всех подстрок длины L и сложим их в сет. Далее будем перебирать подстроки только одной строки, например, первой, и искать хэши в сетах для других строк. Если хэш выбранной подстроки удалось найти во всех сетах, значит, она подходит для ответа - увеличиваем левую границу бинарного поиска, иначе уменьшаем правую. Асимптотика решения $O(NK \log^2 N)$, где N - максимальная длина строки из набора.

2 Частотный анализ

Решение этой задачи предлагается понять самостоятельно. Код можно посмотреть тут <http://pastie.org/5555018>

3 Палиндромы

Подстроку, являющуюся палиндромом, можно представить двумя способами - парой (i, j) , что означает подстроку $s[i..j]$, и парой (i, r) , что означает подстроку $s[i - r..i + r]$ или $s[i - r..i + r + 1]$ в зависимости от четности или нечетности длины палиндрома. Выберем второй способ представления. Зафиксируем центр палиндрома. Пусть мы рассматриваем палиндромы нечетной длины. Палиндромы четной длины рассматриваются аналогично. Нам нужно найти такое наибольшее число r , что подстрока $s[i - r..i - 1]$ совпадает с подстрокой $s[i + 1..i + r]$. Сделаем бинарный поиск по ответу и будем сравнивать хэши подстрок для проверки совпадения. Когда мы нашли нужное r , заметим, что палиндромами также являются подстроки вида $s[i - r + 1..i + r - 1]$, $s[i - r + 2..i + r - 2]$ и так далее. Значит, к ответу нужно прибавить r . Затем нужно так же рассмотреть случай палиндромов четной длины.

4 Плохое хеширование

Заметим, что в качестве ответа мы можем вывести s и $'a' + s$, где s - произвольная строка длины 32. Почему это верно? Во второй строке слагаемое, отвечающее за первую букву 'a', будет иметь вид p^{32} , а поскольку p четно, то его можно представить в виде $p = 2q$. Значит, $(2q)^{32} = 2^{32} \cdot q^{32} = 0(mod 2^{32})$. Поэтому слагаемое не вносит никакого вклада в хэш строки и хэши строк s и $'a' + s$ будут одинаковы.