

Отчёт по алгоритму Рабина-Карпа

Rolling hash

Rolling hash — это способ быстро пересчитать хеш подстроки при сдвиге окна на один символ. Хеш подстроки длины M вычисляется как полиномиальный хеш:

$$\text{hash}(s) = s_0 \cdot p^{M-1} + s_1 \cdot p^{M-2} + \dots + s_{M-1} \cdot p^0 \pmod{mod},$$

где s_0, s_1, \dots — числовые коды символов, p — база (например, 31), а mod — большое простое число для ограничения размера хеша.

Для пересчёта хеша при сдвиге окна на один символ вправо используют формулу:

$$\text{new_hash} = (\text{old_hash} - \text{old_char} \cdot p^{M-1}) \cdot p + \text{new_char} \pmod{mod}.$$

После совпадения хеша выполняется посимвольная проверка, чтобы исключить коллизии.

Оценка сложности

- Вычисление хеша подстроки pattern: $O(M)$
- Вычисление хеша первой подстроки текста: $O(M)$
- Пересчёт rolling hash для всех окон текста: $O(N)$
- Посимвольная проверка при совпадении хеша: $O(M)$ в худшем случае

Средняя сложность: $O(N) + O(M) + O(M) = O(N + M)$

Худшая сложность: $O(N \cdot M)$ (когда много коллизий)

Память: $O(1)$