

Алгоритм Рабина-Карпа

11.12.2024

Богданова Арина

Алгоритм Рабина-Карпа

- Один из алгоритмов поиска подстроки длиной m в строке длиной n
- Редко используется для поиска одиночного шаблона, но эффективен в поиске совпадений множественных шаблонов одинаково длины
- Использует хеширование
- Сильно зависит от качества хеш-функции:
 - В лучшем случае выполняется за $O(n)$
 - В худшем – за $O(nm)$
- Применим в антиплагиате

Идея алгоритма

- Хешируем образец (его длина m)
- Хешируем кусок строки длиной m
- Сравниваем
- Равны?:
 - Да, - сравниваем посимвольно
 - Нет, - идем дальше

a ⁵	a ¹⁰	b ²¹	a	b	a	b
b ²¹	a	b	a	b		
	b ²¹	a	b	a	b	
		b ²¹	a	b	a	b

a ¹⁰	b ²¹	a	b	a	
	a ¹⁰	b ²¹	a	b	a

Хеш-функция

- Если хеш-функция присваивает числа каждому символу и возвращает сумму - крах
- Выход – полином:
 - Возведение в степень:
 - Схема Горнера

$$H(P) = p[0] * x^{m-1} + p[1] * x^{m-2} + \dots + p[m-1] * x^0$$

$$H(P) = p[m-1] * x^{m-1} + p[m-2] * x^{m-2} + \dots + p[0] * x^0 = (\dots (p[m-1] * x + p[m-2]) * x + \dots + p[1]) * x + p[0]$$

- Для реализации подойдет любая кольцевая хеш-функция
- В оригинале – полиномиальная хеш-функция

a ⁵	a ¹⁰	b ²¹	a	b	a	b
b ²¹	a	b	a	b		
	b ²¹	a	b	a	b	
		b ²¹	a	b	a	b