Алгоритмы и модели вычислений.
Поиск подматрицы

Сергей Володин, 272 гр.

16 мая 2014 г.

1 Определения

1. Даны матрицы $P\colon m \times n$, $T\colon M \times N$, состоящие из 1 и 2. $M > m$, $N > n$.
2. Нумеруем строки и столбцы с нуля.
3. Определение: P входит в T с позиции $(i, j) \Leftrightarrow \forall (a, b) \in \overline{0, m-1} \times \overline{0, n-1} \hookrightarrow p_b^a = t_{j+b}^{i+a}$
4. Задача: найти пары (i, j), такие что P входит в T с (i, j).

2 Сведение к поиску подстроки

1. Определим $r(i) = \lfloor \frac{i}{N} \rfloor$ — номер строки, $c(i) = i \mod N$ — номер столбца (см. далее).
2. Запишем матрицу T в видем массива (строки A) $t_i = t_{c(i)}^{r(i)}$. $|t| = MN$.
3. Пусть $p^0, ..., p^{m-1}$ — строки матрицы P, записанные как массив чисел (строки). Обозначим $p = p^0 0^{N-n} p^1 0^{N-n} ... p^{m-1}$. $|p| = Nm + n$. Тогда $p_i = p_{c(i)}^{r(i)}$
4. Пусть p входит в t с позиции i, причем $c(i) \leqslant N - n$, $r(i) \leqslant M - m$. Докажем, что это равносильно P входит в T с позиции $(r(i), c(i))$:

   (a) p входит в t с позиции i $\boxed{\Leftrightarrow}$ $\forall k \in \overline{0, Nm+n-1} \hookrightarrow \left[\begin{array}{lcl} p_k & = & t_{k+i} \\ p_k & = & 0 \\ t_{k+i} & = & 0 \end{array} \right.$. В t нет символов 0, поэтому $\boxed{\Leftrightarrow} \forall k \in \overline{0, Nm+n-1} \hookrightarrow \left[\begin{array}{lcll} p_k & = & t_{k+i} & (1) \\ p_k & = & 0 & (2) \end{array} \right.$. При $c(k) \in \overline{0, n-1}$ $p_k \neq 0$, поэтому должно выполняться (1), для остальных должно выполняться (2), так как t не содержит 0. Поэтому $\boxed{\Leftrightarrow} \forall k \in \overline{0, Nm+n-1}\colon c(k) \in \overline{0, n-1} \hookrightarrow p_k = t_{k+i} \Leftrightarrow \forall k \in \overline{0, Nm+n-1}\colon c(k) \in \overline{0, n-1} \hookrightarrow p_{c(k)}^{r(k)} = t_{c(k+i)}^{r(k+i)}$

3 Время работы

1. $|t| = MN$, $|p| = Nm+n$, поэтому алгоритм поиска подстроки работает за $O(MN \log MN)$ (доказано в решении задания).
2. Заметим, что «очевидная» версия (перебор) работает за $O(MNmn)$ — $O(MN)$ позиций и $O(mn)$ на проверку вхождения.