1 Алгоритм решения

1. P разбивается символами «?» на k непустых подстрок $P_1, P_2, \dots, P_k$. Для каждой подстроки P_i сохраняем её смещение $shift_i$ от конца образца, т.е. количество символов (включая «?») между концом P_i и концом P.

2. Построение бора и суффиксных ссылок на $\{P_1, \dots, P_k\}$. Делаем **Ахо–Корасик**.

3. Каждая вершины бора - конец подстроки, держит индекс подстроки в массиве $\{P_1, \dots, P_k\}$.

4. Массив счётчиков $Count[0 \dots |T| - 1] := 0$, где $Count[i]$ - сколько из k подстрок совпали на позиции i текста.

5. Проходим T по бору. Если при индекс текущего символа в тексте j вершина соответствует концу P_i, увеличиваем $Count[j + shift_i]$ на 1.

6. Итоговые индексы i, т.ч. $Count[i - |P| + 1] = k$ и $i - |P| + 1 \geqslant 0$, соответствуют позициям начала вхождений образца P в текст T.

При этом, достаточно кольцевого буфера длины $|P|$ заместо массива $Count$ длины $|T|$, поскольку на каждой итерации алгоритм обновляет только одну позицию, соответствующую текущему окну длины $|P|$.

2 Корректность

Теорема 1. *$T[i \dots i + |P| - 1]$ является вхождением образца P, если и только если:*

1. *Для каждой подстроки P_j существует точное вхождение P_j в T на позиции $i + |P| - shift_j - |P_j|$.*

2. *Символы $T[i \dots i + |P| - 1]$, не покрытые подстроками P_j являются «?».*

Доказательство. **Необходимость.** Пусть подстрока $S = T[i \dots i + |P| - 1]$ действительно совпадает с образцом P, учитывая, что символы «?» могут принимать любые значения. Тогда каждая фиксированная часть P_j должна найти точное вхождение в S на отрезке

$$i + |P| - shift_j - |P_j| \ \dots \ i + |P| - shift_j - 1,$$

что эквивалентно точному совпадению P_j в тексте T на позиции $i + |P| - shift_j - |P_j|$. Остальные позиции в S соответствуют символам «?» и поэтому не накладывают дополнительных ограничений.

Достаточность. Предположим теперь, что для каждого $j = 1, \dots, k$ имеется точное совпадение P_j в указанных выше позициях, а все остальные символы в S могут быть произвольными. Тогда из определения символа «?» следует, что эти «нежёсткие» позиции не препятствуют полному совпадению S с P. Следовательно, S является вхождением P в текст. □

Таким образом, если $Count[i] = k$, то это означает, что каждая подстрока P_j сопоставилась с нужной позицией относительно начала вхождения, а значит, P сопоставим с $T[i \dots i + |P| - 1]$.

3 Временная сложность

Построение бора и суффиксных ссылок для k строк суммарной длины $O(|P|)$ выходит в $O(|P|)$, поиск в тексте T осуществляется за $O(|T| + |R|)$, где $|R|$ - длина ответа поиска, которое зависит от общего количество найденных вхождений в тексте, максимальное значение которого ограничевается количеством всех образцов на каждый символ текста. Количество образцов зависит напрямую зависит от количества символов «?», допустим $|Q|$, откуда $\max |R| = |T|(|Q| + 1)$, из чего Ахо-Корасик имеет сложность $O(|T||Q|)$ $\to O(|P| + |T||Q|)$ - общая сложность по времени.

4 Затраты по памяти

Бор - $O(|P|)$, кольцевой буфер - $O(|P|)$, структуры данных Ахо–Корасика - $O(|P|) \to O(|P|)$ - итоговая пространственная сложность алгоритма.