



## pattern – Soluție

Autor, ing. Silviu Găncăanu

Problema se poate rezolva prin programare dinamică în modul următor:

$T(i, j)$  = numărul de posibilități de potrivire a primelor  $j$  caractere din mască peste o subsecvență din text care se termină la poziția  $i$ . Recurența este următoarea:

$T(i, j) =$

- 0, dacă mască( $j$ ) este un caracter diferit de text( $i$ )
- $T(i-1, j-1)$  dacă mască( $j$ ) = text( $i$ ) sau mască( $j$ ) = ?
- suma( $T(k, j-1)$ ) pentru  $k$  de la 1 la  $i$  pentru mască( $j$ ) = \*

Dacă întreținem și sumele parțiale, obținem o complexitate de  $O(\text{lungime}(\text{text}) * \text{lungime}(\text{mască}))$  pentru fiecare test din fișierul de intrare. Această soluție implementată fără a ține toată matricea (pentru a intra în limita de memorie de 8 MB) obține ~40% din punctaj.

Soluția de mai sus nu valorifică faptul că sunt maxim 100 caractere speciale în mască. Pentru a folosi acest lucru spargem masca în num\_wildcards+1 bucăți (num\_wildcards = numărul de caractere speciale). Aflăm locurile din text unde bucățile se potrivesc cu un algoritm liniar de potrivire a șirurilor (KMP, Rabin-Karp, etc.). Complexitatea pentru toate bucățile va fi

$O(\text{num\_wildcards} * \text{lungime}(\text{text}) + \sum(\text{lungime}(\text{bucata}_i))) =$   
 $O(\text{num\_wildcards} * \text{lungime}(\text{text}) + \text{lungime}(\text{mască})).$

Acum putem rezolva problema tot prin programare dinamică, dar mai eficient. Fie  $\text{Cnt}(i, j)$  = numărul de posibilități de a potrivi caracterelor din mască până la al  $j$ -lea caracter special peste o subsecvență din text care se termină la poziția  $i$ . Notăm cu  $S_x$  subsecvența de litere (eventual vidă) din mască dintre caracterele speciale  $x-1$  și  $x$  (inclusiv cele dintre primul caracter special și începutul textului și dintre ultimul caracter și sfârșitul textului) și cu  $P_x$  poziția caracterului special  $x$  în mască. Acum avem recurența:

$\text{Cnt}(i, j) =$

- $\text{Cnt}(i - \text{lungime}(S_j) - 1, j - 1)$  dacă mască( $P_j$ ) = ? și  $S_j$  se potrivește peste subsecvența text( $i - \text{lungime}(S_j) : i - 1$ )
- $\sum(\text{Cnt}(i - \text{lungime}(S_j) - k, j - 1))$  pentru fiecare  $k$  de la 0 la  $i - \text{lungime}(S_j)$  pentru care avem că  $S_j$  se potrivește peste subsecvența text( $i - \text{lungime}(S_j) - k - 1 : i - k$ ), dacă mască( $P_j$ ) = \*

Recurența se poate implementa eficient știind locurile unde subsecvența  $S_j$  se potrivește în text.

Complexitatea pentru recurența este  $O(\text{lungime}(\text{text}) * \text{num\_wildcards})$  iar complexitatea în total este  $O(\text{lungime}(\text{text}) * \text{num\_wildcards} + \text{lungime}(\text{mască}))$ .

Cazurile în care masca începe și/sau se termina cu \* sau ? trebuie tratate cu atenție. Initializările sau modul de calcul final al modalităților de potrivire va fi diferit.