

文字列照合アルゴリズムの時間計算量は、テキスト中の文字の種類数が C であるとき、最良時間計算量が $O\left(C + \frac{n}{m}\right)$ 、最悪時間計算量が $O(C + n)$ となる。

演習問題

12.1 以下の文章の①～⑦について、それぞれ正しい記号を下から選べ。正しい記号が複数存在する場合はすべて列挙せよ。ただし、②、③、⑥、⑦については、もっとも適切なものを1つだけ選ぶこと。

文字列照合を行う基本的なアルゴリズムは、(①)するが、長さ n のテキストと長さ m のパターンを入力とする場合、パターンがテキスト中に存在しなければ、その最良時間計算量は(②)であり、最悪時間計算量は(③)である。

ボイヤー-ムーア法とよばれる文字列照合アルゴリズムは、1つ目のアイデアにより、(④)するが、これにより不一致が起こった場合にパターンを最大(⑤)文字分だけ右にずらしてつぎの比較を行うことができる。また、2つ目のアイデアも同時に用いて、テキスト中の文字の種類数を定数と考えると、パターンがテキスト中に存在しない場合のボイヤー-ムーア法の最良時間計算量は(⑥)であり、最悪時間計算量は(⑦)である。

- ① : a. テキストとパターンをパターンの左端から比較
b. テキストとパターンをパターンの右端から比較
c. テキストの右端から比較を開始 d. テキストの左端から比較を開始
- ② : a. $O(mn)$ b. $O(n + m)$ c. $O\left(\frac{n}{m}\right)$ d. $O(n)$
- ③ : a. $O(mn)$ b. $O(m)$ c. $O\left(\frac{n}{m}\right)$ d. $O(n)$
- ④ : a. テキストとパターンをパターンの左端から比較
b. テキストとパターンをパターンの右端から比較
c. 不一致が起こったテキスト中の文字の情報を利用
d. 不一致が起こったパターン中の文字の情報を利用
- ⑤ : a. m b. n c. $\frac{n}{m}$ d. mn
- ⑥ : a. $O(mn)$ b. $O(n + m)$ c. $O\left(\frac{n}{m}\right)$ d. $O(n)$
- ⑦ : a. $O(mn)$ b. $O(m)$ c. $O\left(\frac{n}{m}\right)$ d. $O(n)$

12.2 “therefore” というパターン $P[0] \sim P[8]$ に対して、ホールスプールのアルゴリズムで用いられる配列 $S[t], S[h], S[e], S[r], S[f], S[o]$ の値を示せ。

12.3 “therefore” というパターン $P[0] \sim P[8]$ に対して、ボイヤー-ムーア法の1つ目のアイデアで用いられる配列 $L[t], L[h], L[e], L[r], L[f], L[o]$ の値を示せ。

12.4 “therefore” というパターン $P[0] \sim P[8]$ に対して、ボイヤー-ムーア法の2つ目のアイデアで用いられる配列 $S[0] \sim S[8]$ の値を示せ。

12.5 “A man said, “I think, therefore I am.”” というテキスト(最初の A からピリオドの後ろのダブルクオーテーションまでがテキストで、ダブルクオーテーション、空白、コンマなども 1 文字と数えるものとする)と“therefore”というパターンを入力として、ポイヤームー法の 2 つのアイデアを同時に用いて文字列照合を行うとき、テキストとパターンの文字の比較は何回実行されるかを答えよ。