UNIVERSITÉ DE ROUEN
UFR des Sciences
et des Techniques

MASTER PREMIÈRE ANNÉE
G.I.L. et I.T.A.
Année 2010-2011

Algorithmique du texte

Contrôle continu — 3 décembre 2010

Support de cours et notes de TD autorisés. Calculatrices, mobiles et portables interdits. Durée : 1 h 30. Le barème est donné à titre indicatif. Les exercices III et IV sont au choix.

Exercice I *(5 points.)*

En vous appuyant sur une propriété que vous commencerez par citer :

1) Quelle est la forme des plus longs mots qui admettent 2 et 3 comme périodes mais pas 1 ?

2) Même question mais avec 5 et 6 en lieu et place de 2 et 3.

3) De manière générale maintenant, quelle est la forme des plus longs mots qui admettent p et $p+1$ comme périodes mais pas 1 ? (Vous supposerez acquis que le plus grand diviseur commun à n et $n+1$ est égal à 1 pour tout naturel non nul n.)

Exercice II *(5 points.)*

Pour le mot $x = \mathtt{abacaababac}$:

1) Dressez :

a) la table du bon préfixe ;

b) la table du meilleur préfixe.

2) Dessinez :

a) la version automate implanté par fonction de suppléance et successeur par défaut de l'algorithme de Morris et de Pratt ;

b) même chose pour l'algorithme de Knuth, de Morris et de Pratt ;

c) l'automate de Simon.

Exercice III *(10 points.)*

On cherche, une fois de plus, à montrer que les algorithmes « naïfs » de recherche d'un mot x non vide dans un texte y effectuent un nombre raisonnable de comparaisons de lettres de x et de y dans le pire des cas pour un nombre non négligeable de mots x.

On s'intéresse ici :

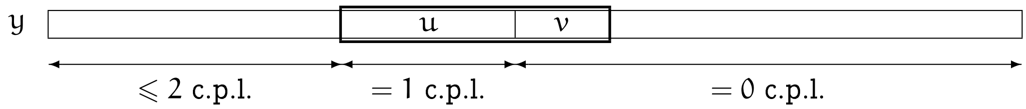
— à l'algorithme naïf qui effectue la comparaison des lettres du mot x du facteur de y délimité par la fenêtre glissante de taille $|x|$ de la gauche vers la droite (en croissant de la position 0 vers la position $|x|-1$ conjointement sur x et sur le facteur donc) ;

— aux mots x tels que $|x| \geqslant 2$ et que la lettre $x[0]$ n'apparait à nulle autre position que la position 0 dans x ;

— aux textes y tels que $|y| \geqslant |x|$.

1) Sous l'hypothèse d'une distribution uniforme et indépendante des lettres de l'alphabet, quelle est, en fonction du cardinal c de l'alphabet, la probabilité des mots x considérés ? *Remarque :* une probabilité proche de 1 rend pertinente l'analyse proposée ici ; une probabilité proche de 0 la rend insignifiante.

2) Montrez que chaque lettre du texte y est comparée au plus deux fois à des lettres du mot x. *Suggestion :* dessinez et précisez l'invariant de la boucle principale de l'algorithme ; celui-ci pourrait avoir la forme :



> *où* uv *est le contenu de la fenêtre…* u *est de longueur…* u *est de la forme… les lettres de* y *positionnées avant la fenêtre ont été comparées au plus deux fois et celles situées après la fenêtre aucune fois… les lettres de* u *ont été comparées une fois et une seule… les lettres de* v *n'ont jamais été comparées… c.p.l. signifie « comparaison(s) par lettre »…*

le seul dessin avec ses précisions ne suffit évidemment pas : il faut également montrer qu'il s'agit bien d'un invariant de boucle.

3) Déduisez-en que le nombre de comparaisons de lettres du mot x et du texte y est majoré par $2|y|$.

4) Donnez un exemple de mot x et de texte y (de longueurs non constantes) pour lequel la borne $2|y|$ est presque atteinte.

Exercice IV *(10 points.)*

1) Montrez que si deux mots x et y sont tels que $x^m = y^n$ pour deux naturels non nuls m et n, x et y sont des puissances d'un même mot z. *Suggestion :* lorsque nécessaire, appliquer le lemme de périodicité au mot $t = x^m = y^n$.

Rappelons qu'un mot non vide est *primitif* s'il n'est puissance d'aucun autre mot que lui-même.

2) Donnez la forme — ou les formes le cas échéant — des mots de longueur 6 qui ne sont pas primitifs.

3) Montrez que pour tout mot non vide x, il existe un et un seul mot primitif dont x est une puissance.