M1 Bio-Info, Info et Math-Info
Algorithmique du texte
HMSN116M

Année scolaire 2016-2017

TD 1 : Algorithmique du texte

Sèverine Bérard - Université de Montpellier, ISE-M

Severine.Berard@umontpellier.fr

1 Pour s'échauffer ...

Exercice 1 *Quel est l'alphabet de taille minimum qui permette d'écrire les mots* `allo`, `logo`, `algo` *et* `galop` *?*

Exercice 2 *Mots d'un alphabet :*

1. *Quels sont tous les mots de taille 3 que l'on peut écrire avec l'alphabet* $\{0, 1\}$ *?*
2. *Plus généralement, combien de mots de taille* k *peut-on écrire avec un alphabet de taille* n *?*
3. *Combien de mots en tout peut-on écrire avec un alphabet de taille* n *?*

Exercice 3 *Devinette :*

1. *Peut-on déterminer* u *et* v*, deux mots de l'alphabet* $\{a, b, c\}$ *sachant que :*
   — $|u| = |v| + 2$
   — $|u|_a = 2 * |v|_a$
   — $|u|_b = |v|_b = 1$
   — $|u.v| = 4$
2. *Et si on ajoute l'information que le mot* `cc` *n'est ni préfixe, ni suffixe de* u *?*

2 À partir des mots

Exercice 4 (Chevauchement) *Deux mots* se chevauchent *si on peut les écrire l'un sous l'autre de manière à ce que les lettres se situant dans une même colonne soient identiques. Par exemple, les mots* `010011` *et* `1100101` *se chevauchent, tout comme les mots* `mobilité` *et* `playmobil`*. Ces deux chevauchements sont* non triviaux*. Les mots* `abcaacbacc` *et* `bca` *se chevauchent de manière triviale, la partie commune étant un des mots tout entier :*

```
010011                 mobilité     abcaacbacc
   1100101          playmobil        bca
```

En fonction des notions de préfixe, suffixe et/ou facteur que vous avez vues en cours, exprimez :

1. *la notion de chevauchement non trivial entre deux mots* u *et* v
2. *la notion de chevauchement entre deux mots* u *et* v

Exercice 5 *Soit* Σ *l'alphabet français à 26 lettres.*

1. *Trouvez un mot* $u \in \Sigma^*$ *tel que* `ent` *est un suffixe de* u *et* `aiet` *en est une sous-séquence.*
2. *Trouvez le mot* $v \in \Sigma^*$ *de longueur minimale répondant aux critères ci-dessus.*
3. *Trouvez un mot* $w \in \Sigma^*$ *tel que* `ami` *est un préfixe de* w *et* `iti` *en est un sous-mot.*
4. *Trouvez le mot* $x \in \Sigma^*$ *de longueur minimale répondant aux critères ci-dessus.*
5. *Trouvez un mot* $y \in \Sigma^*$ *tel que* `iment` *est un suffixe de* y*,* `infi` *en est un préfixe et* `ii` *en est une sous-séquence.*
6. *Trouvez le mot* $z \in \Sigma^*$ *de longueur minimale répondant aux critères ci-dessus.*
7. *Trouvez un mot* $a \in \Sigma^*$ *tel que* `grb`*,* `iou` *et* `roi` *en soient des sous-séquences.*
8. *Trouvez un mot* $b \in \Sigma^*$ *de longueur minimale répondant aux critères ci-dessus. Y a-t-il plusieurs mots* b *possibles ?*

Exercice 6 (Lemme de Lévi, 1994 (extrait)) *Pour tous mots* $x, y, z, t \in A^*$, $x.y = z.t$ *implique qu'il existe un mot* $u \in A^*$ *tel que*

— *soit* $x = z.u$ *et* $t = u.y$

— *soit* $z = x.u$ *et* $y = u.t$

Donnez une preuve de ce lemme, on pourra utiliser une preuve graphique.

3 Bords et périodes

On rappelle la proposition sur la caractérisation des périodes vue en cours :

Proposition 1 *Soient* x *un mot non vide et* p *un entier tel que* $0 < p \leqslant |x|$. *Alors les 5 propriétés suivantes sont équivalentes :*

1. *L'entier* p *est une période de* x
2. *Il existe deux mots* u *et* $v \neq \epsilon$ *et un entier* $k > 0$ *tels que* $x = (uv)^k u$ *avec* $|uv| = p$
3. *Il existe un mot* t *et un entier* $k > 0$ *tels que* x *est un préfixe de* t^k *avec* $|t| = p$
4. *Il existe trois mots* u, v *et* w *tels que* $x = uw = wv$ *et* $|u| = |v| = p$.
5. *Il existe un mot* t *tel que* x *est préfixe de* tx *et* $|t| = p$

Exercice 7 *Démontrez la proposition ci-dessus.*

Exercice 8 *On rappelle qu'un mot* w *est dit* sans bord *si son seul bord est le mot vide, c'est-à-dire si* $period(w) = |w|$. *On suppose qu'un mot* x *a un bord de longueur minimale, non vide,* u. *Montrer que* u *est sans bord et qu'il existe un mot* v *tel que* $x = uvu$.

Exercice 9 *Soit* w *un mot non vide,* u *un bord de* w, *avec* $u \neq \epsilon$, *et* v *un bord de* u, *avec* $v \neq \epsilon$.
Est-ce que la propriété suivante est vraie : $|w| > |u| + |v| + 1$ *? Si oui, la prouver, sinon donnez un contre-exemple.*

Exercice 10 (Ordre lexicographique) *On considère dans cet exercice l'ordre lexicographique, c'est-à-dire "l'ordre du dictionnaire", on le note* $>_{\mathcal{L}}$. *On a donc par exemple* `chien` $>_{\mathcal{L}}$ `chat` $>_{\mathcal{L}}$ `ch`.
Pour un mot non vide x, *on définit* $Maxsuf(x)$ *comme étant le suffixe maximal de* x *au sens de l'ordre lexicographique.*

1. *Donner* $Maxsuf(x)$ *pour* $x =$`abacabbac`.
2. *Soit* x *un mot non vide, et* λ *une lettre quelconque. On pose* $u = Maxsuf(x)$ *et on définit* v *tel que* $v\lambda = Maxsuf(x\lambda)$ [1]. *Montrer que soit* $v = u$, *soit* v *est un bord de* u.

On rappelle le Lemme de périodicité faible :

Lemme 1 (Périodicité faible) *Soient* p *et* q *deux périodes d'un mot* x. *Si* $p + q \leq |x|$, *alors* $pgcd(p, q)$ *est aussi une période de* x.

Exercice 11 *On veut démontrer ce lemme.*

1. *Traiter le cas* $q = p$.
2. *Si* $p > q$, *montrer que* $pgcd(p, q) = pgcd(p - q, q)$.
3. *Montrer que si* $p > q$ *et* $p + q \leq |x|$ *alors* $p - q$ *est une période de* x.
4. *Conclure en réalisant une récurrence sur* $\max(p, q)$.

Exercice 12

1. *Soit* x *un mot non vide. Soit* u *le plus petit mot tel que* x *est préfixe de* ux. *Montrer que* $|u| = period(x)$.
2. *Soit* x *un mot non vide. Montrer que les trois propositions suivantes sont équivalentes :*
   (a) $period(x^2) = |x|$,
   (b) x *est primitif, c'est-à-dire ne peut pas être écrit sous la forme* u^k *pour* $k > 1$,
   (c) x^2 *contient seulement 2 occurrences de* x.

1. $v\lambda$ est la concaténation du mot v avec la lettre λ, de même $x\lambda$ est la concaténation du mot x avec la lettre λ.