Colle n^o 4 d'Informatique – CaML
Sous-mot et facteur d'un mot

Mathieu Goutelle (Mathieu.Goutelle@ens-lyon.fr)

5 – 9 janvier 2004

1 Définitions

On appelle **sous-mot** d'un mot v un mot u obtenu en supprimant zéro, une ou plusieurs lettres du mot v mais en conservant l'ordre initial des autres lettres. En termes plus mathématiques, un mot u de longueur p est un sous-mot du mot v de longueur n s'il existe une application strictement croissante s tel que : $\forall\ 1 \leq j \leq p, u_j = v_{s(j)}$. On convient que ε est un sous-mot de tout mot v.

On appelle **facteur** d'un mot v un mot u apparaissant au moins une fois dans v. On peut écrire cela ainsi : u est un facteur de v s'il existe deux mots x et y tel que $v = xuy$.

Remarque : Ce que j'appelle ici **facteur** est appelé par les anglophones *subword* et ce que j'appelle ici **sous-mot** est appelé *subsequence* ou *scattered subword* (parfois *pattern*).

2 Préliminaires

2.1 Écrire une fonction `list_of_string` qui transforme une chaîne de caractéres en une liste de caractères.

2.2 Écrire une fonction `string_of_list` qui réalise l'opération inverse de la précédente.

3 Recherche de sous-mot

Programmer la recherche d'un sous-mot dans un mot.

4 Recherche de facteur

4.1 Programmer la recherche d'un facteur.

4.2 Faire fonctionner votre fonction « à la main » sur le mot `"banananobano"` et le préfixe `"nano"`. Calculer la complexité de votre fonction.

4.3 Montrer que certaines comparaisons sont inutiles et proposer une idée pour améliorer le nombre de comparaisons. *Pour les curieux, des algorithmes efficaces pour une telle recherche sont par exemple les algorithmes de Knuth-Morris-Pratt et Boyer-Moore.*

Bonne Année à tous !