Devoir à la maison
Théorie des Langages (INF105)

Pierre SENELLART
pierre.senellart@telecom-paristech.fr

1 Numéros de téléphone (12 points)

On souhaite caractériser le langage des numéros de téléphone afin de construire un outil d'extraction automatique depuis des emails ou pages Web. On fait les hypothèses suivantes :

— Un numéro de téléphone est une suite non vide de chiffres décimaux, éventuellement séparés en groupes non vides à l'aide d'un ou plusieurs traits d'unions (p. ex., « 1234567 », « 123-456-789 », « 12-24-36 », « 1234-5678 », mais ni « 12--34 » ni « -1234567 »).

— Un numéro de téléphone peut être précédé d'un indicatif international (une suite non vide de chiffres décimaux), qui peut être indiqué de l'une des trois manières suivantes : « (+44) », « +44 », « (44) ». Ainsi, « (+25)13456789 »ou « +1313-45-27 » sont des numéros de téléphone valides. L'indicatif téléphonique n'est pas obligatoire.

1. (2 points) Les hypothèses ci-dessus sont simplificatrices. Donner deux exemples (de types différents) de numéros de téléphone du monde réel qui ne respectent pas ces règles ; inversement, donner deux exemples (de types différents) de mots respectant les règles ci-dessus qui ne sont pas des numéros de téléphones du monde réel.

   Dans ce qui suit, on oubliera ces cas particuliers et **on supposera que les numéros de téléphones à traiter sont ceux définis par les hypothèses données plus haut**.
2. (1 point) Indiquer l'alphabet sur lequel sont définis les numéros de téléphones.
3. (2 points) Donner une expression rationnelle qui décrit le langage des numéros de téléphone.
4. (3 points) En utilisant une construction vue en cours, éventuellement simplifiée, donner un automate fini qui reconnaît le même langage. Indiquer les étapes de votre construction.
5. (2 points) Déterminiser cet automate, s'il n'est pas déjà déterministe.
6. (2 points) Minimiser l'automate obtenu ; indiquer les étapes de la procédure.

2 Clôture par intersection infinie (8 points)

Soit l'alphabet $\Sigma = \{a, b\}$.

On dit qu'un ensemble de langages $\mathcal{L} \subseteq 2^{\Sigma^*}$ est *clos par intersection infinie* si pour toute suite $(L_n)_{n\in\mathbb{N}}$ de langages de $\mathcal{L}$, $\bigcap_{n\in\mathbb{N}} L_n \in \mathcal{L}$. Le but de cet exercice est d'étudier la clôture des rationnels par intersection infinie.

On dit qu'un mot u de Σ^* est le *transposé* (ou *miroir*) d'un mot $v = v_1 \dots v_n$ (noté $u = \bar{v}$) si u s'écrit $v_n \dots v_1$. Ainsi, le transposé de *abbbbaba* est *ababbbba*. Un mot u est un palindrome s'il est identique à son transposé : $u = \bar{u}$. Pour un entier n donné, le préfixe (respectivement, suffixe) de longueur n d'un mot u est le sous-mot formé des n premiers (respectivement, derniers) symboles de u. Ainsi, le préfixe de longueur 3 de *abbbbaba* est *abb* tandis que son suffixe de longueur 3 est *aba*. Nous nous intéressons à la suite de langages L_n ($n \geqslant 1$) sur l'alphabet Σ, définie de la manière suivante. Pour tout entier $n \geqslant 1$, L_n est l'ensemble des mots de longueur supérieure ou égale à $2n$ dont le suffixe de longueur n est le transposé du préfixe de longueur n. Ainsi, $abbbbaba \in L_1 \cap L_2$ (car a est le transposé de a et ab le transposé de ba) mais $abbbbaba \notin L_3$.

1. (1 point) Donner une expression rationnelle décrivant le langage L_1, formé des mots de longueur supérieure ou égale à 2 dont le premier symbole est identique au dernier symbole.
2. (1 point) Donner une expression rationnelle décrivant le langage L_2.
3. (2 points) En s'inspirant des questions précédentes, montrer que L_n est un langage rationnel pour tout $n \geqslant 1$.
4. (2 points) On considère maintenant, pour $n \geqslant 1$, le langage L'_n formé de tous les mots de longueur strictement inférieure à $2n$, ainsi que des mots de longueur supérieure ou égale à $2n$ dont le suffixe de longueur n est le transposé du préfixe de longueur n. Pour $n \geqslant 1$, L'_n est-il un langage rationnel ? Prouver votre affirmation.
5. (2 points) En déduire que les langages rationnels ne sont pas clos par intersection infinie. On admettra sans démonstration que le langage des palindromes n'est pas un langage rationnel.