L2 MIASHS, Université de Lorraine

Ingénierie linguistique : TD 4

Chuyuan Li, Marie Cousin

Feb. 03, 2023

1 Reprise du TD3 - Expressions régulières 2

1. On applique l'algorithme de recherche des chaînes de caractères dans un texte correspondant à une E.R. donnée. Dans chacun des cas, donner le résultat de la recherche:

   1.1 E.R.: "a\$w^*$a" Texte: "acddcceeacvvv bbaa"

   1.2 E.R.: "[^ a-z]^+ba" Texte: "Aaa Abbca bac"

   1.3 E.R.: "^ .\$b$" Texte: "Je viens.\$n$ Et toi?"

   1.4 E.R.: "\b[A-Z].*" Texte: "Tu connais Paul et Marie ?"

   1.5 E.R.: "[\s .] [A-Z]" Texte: "Tu connais l'O.N.U. C'est difficile."

2. Appliquer l'algorithme de recherche de la question 1 pour les E.R. suivantes dans le texte ci-dessous[1] : Parmi l'ensemble des expressions renvoyées, en donner 3 pour chacune des E.R. Si dessous. Sinon, justifier pourquoi cela n'est pas possible.

   À LÉON WERTH

   Je demande pardon aux enfants d'avoir dédié ce livre à une grande personne. J'ai une excuse sérieuse : cette grande personne est le meilleur ami que j'ai au monde. J'ai une autre excuse : cette grande personne peut tout comprendre, même les livres pour enfants. J'ai une troisième excuse : cette grande personne habite la France où elle a faim et froid. Elle a bien besoin d'être consolée. Si toutes ces excuses ne suffisent pas, je veux bien dédier ce livre à l'enfant qu'a été autrefois cette grande personne. Toutes les grandes personnes ont d'abord été des enfants. (Mais peu d'entre elles s'en souviennent.) Je corrige donc ma dédicace :

   À LÉON WERTH
   QUAND IL ÉTAIT PETIT GARÇON

   2.1 E.R.: "a\$w^*$e"

   2.2 E.R.: "[^ A-Z]^+nt"

   2.3 E.R.: ".\$b$$"

   2.4 E.R.: "\b[A-Z].$^?$"

[1] Dédicace du "Petit Prince" d'Antoine de Saint-Exupéry, disponible ici

2 Reprise du TD 3 - Expressions régulières 3

Définir en français les expressions régulières suivantes

1. $\backslash b \backslash W \backslash b$
2. $\hat{} \backslash W + \$$
3. $\backslash bv.\{4\} \backslash b$
4. $\hat{} v[a-z]\{4,\} \backslash b$

3 Vice-Versa

Donner les expressions régulières pour les descriptions suivantes

1. Deux caractères présents au moins une fois, et pris dans l'ensemble contenant "A" ou "B" ou "C".
2. Toute suite d'au moins deux caractères ne contenant que des paires de caractères identiques pris parmi l'ensemble contanant "A", "B" et "C", à savoir soit AA, soit BB, soit CC.
3. Tout mot d'au moins un caractère pris entre les lettres "d" et "t". Ainsi, "vendredi" par exemple ne sera pas reconnu, car il contient la lettre "v" qui n'appartient pas à la classe de caractères [d-t].
4. Une séquence de 1 à 3 lettres, suivie d'une séquence de 1 à deux chiffres, suivie de la même séquence de chiffres suivie de la première séquence de lettres.