

Ingénierie linguistique : TD 5

Chuyuan Li, Marie Cousin

Feb. 10, 2023

1 Reprise du TD4

1. Appliquer l'algorithme de recherche pour les E.R. suivantes dans le texte ci-dessous¹ : 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

1.1 E.R.: "a\w*e"

1.2 E.R.: "[^A-Z]⁺nt"

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

1.4 E.R.: "\b[A-Z].?"

2. Définir en français les expressions régulières suivantes

2.1 $\backslash bv.\{4\}\backslash b$

2.2 $\wedge v[a - z]\{4, \}\backslash b$

3. Donner les expressions régulières pour les descriptions suivantes

3.1 Deux caractères présents au moins une fois, et pris dans l'ensemble contenant "A" ou "B" ou "C".

3.2 Toute suite d'au moins deux caractères ne contenant que des paires de caractères identiques pris parmi l'ensemble contenant "A", "B" et "C", à savoir soit AA, soit BB, soit CC.

3.3 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].

3.4 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.

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

2 Si déjà fini la section 1

1. Ecrire une expression régulière sur l'alphabet {a, b, c} qui contient
 - une étoile et fini par un a
 - au plus un a
 - au moins 2 a
 - la dixième lettre en partant de la fin est un a
2. Ecrire les expressions régulières dans le cadre général
 - les séquences de 5 caractères dont le retour à la ligne
 - les chaînes de caractères débutant par un nombre arbitraire de \ suivi par des *
 - les nombres en notation scientifique anglaise : au moins un chiffre suivi de :
 - une partie décimale : un point suivi par deux chiffres
 - un exponentiel : un e suivi par une valeur - suivi d'un nombreExemples : 42, 66.4E-5, 8E17, . .

3 Application du Regex

1. Écrire une expression régulière qui s'applique pour tous les exemples suivants :
 - il est bleu, blanc, rouge
 - il est bleu, blanc et rouge
 - Il est bleu, blanc, rouge
 - Il est bleu, blanc et rouge
 - bleu, blanc, rouge
 - bleu, blanc et rouge
2. (Ouverture) Test d'un mot de passe : il doit contenir au moins un chiffre, au moins une lettre majuscule, une lettre minuscule, et au moins un caractère de ponctuation (./\!-_...). De plus il doit avoir entre 8 et 32 caractères.
Exemple : abce.fr/ty2