

Projet DILAF Dictionnaires Langue Africaine – Français



Expressions rationnelles

1 - Expressions rationnelles : localisation

Une expression rationnelle est une suite finie de caractères qui peut servir de modèle pour localiser un ensemble de "mots" précis dans un texte donné. En informatique, ces suites sont appelées, des chaînes de caractères.

Notations

Les expressions rationnelles sont surlignées pour plus de clarté. Il faut faire attention à la lecture de l'expression quand elle comporte des espaces, ceux-ci sont mis en évidence par le symbole "␣". L'expression à tester se fera sans ce symbole mais avec une vrai espace.

Exercices pratiques

Les expressions rationnelles seront utilisées avec l'éditeur de texte Open Office¹.

Utiliser la fonction "Rechercher & Remplacer" de l'éditeur de texte, en allant dans :

Édition > Rechercher & Remplacer

Il existe aussi un raccourci pour ouvrir la fenêtre, appuyer sur les touches 'Ctrl' et 'F' en même temps. Si ce raccourci n'ouvre pas une fenêtre similaire à celle ci-dessous, la version d'Office est peu être antérieur à celle représentée et il faudra essayer avec le raccourci 'Ctrl' et 'H'.

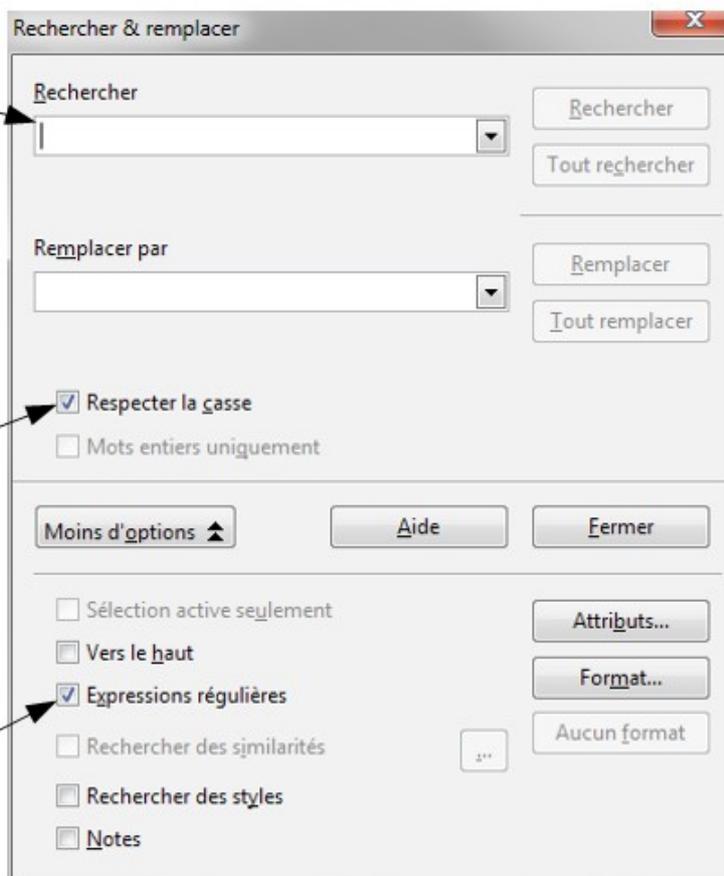
Cet outil sert à la fois pour localiser des expressions, pour remplacer des termes par d'autres, ou les deux. Il n'est pas nécessaire à chaque fois de devoir remplacer l'expression recherchée.

¹ Téléchargeable gratuitement : <http://fr.openoffice.org/>

Placer la chaîne de caractère à localiser dans la case « Rechercher »

Cocher la case "Respecter la casse" afin de distinguer caractères minuscules et majuscules

Cocher la case "Expressions rationnelles" dans les options.



Un bon moyen pour comprendre est de taper les exemples dans la case "Rechercher" et de regarder les résultats obtenus en cliquant sur le bouton "Rechercher" ou "Tout rechercher".

1.1 - Suite de caractères

Recherche d'une chaîne de caractères.

Remarque :

Le symbole de la barre espace du clavier (l'espace) est aussi considéré comme un caractère.

La casse est respectée donc "a" et "A" sont des caractères différents.

Exemple 1.1.1

de

recherche de toutes les chaînes de caractères "de"

Exemple 1.1.2

à

recherche de tout les caractères "à"

1.2 - Disjonction "ou"

La disjonction est notée par le signe "|"².

La disjonction permet de rechercher en même temps plusieurs chaînes de caractères.

Exemple 1.2

`le|la`

recherche de toutes les chaînes de caractères "le" ou "la"

1.3 - Mise en facteur

Les parenthèses permettent de mettre en facteur certaines parties d'une expression.

Exemple 1.3

`(dessus|dessous)_(le|la)`

recherche de toutes les chaînes de caractères "dessus" ou "dessous" suivies d'une espace et de "le" ou "la"

1.4 - Ensemble de caractères

Les crochets "["³ et "]"⁴ permettent de définir un ensemble de caractères recherché.

Exemple 1.4.1

`[aeiouy]`

recherche de toutes les voyelles du texte

Exemple 1.4.2

`[,;.:?!-]`

recherche des signes de ponctuation

Exemple 1.4.3

`[dl][aeiouy]`

recherche de toutes les chaînes de caractères commençant soit par "d" ou "l" suivies d'une voyelle

Exercice 1.4 :

Construire une expression rationnelle qui recherche les mots d'une lettre suivis d'une voyelle élidée comme "d'", "m'", "t'", etc.

2 Clavier azerty : appuyer sur la touche 'Alt Gr' et, en même temps sur la touche '6'.

3 Clavier azerty : appuyer sur la touche 'Alt Gr' et, en même temps sur la touche '5'.

4 Clavier azerty : appuyer sur la touche 'Alt Gr' et, en même temps sur la touche '0'.

1.5 - Intervalle de caractères

La notation entre crochet et avec un tiret ("-") le caractère mathématique moins) permet de spécifier un intervalle de caractères recherchés.

Exemple 1.5.1

`[0-9]`

recherche des chiffres de 0 à 9

Exemple 1.5.2

`[A-Z]`

recherche des lettres majuscules de A à Z

Exemple 1.5.3

`[A-Z0-9]`

recherche des lettres majuscules de A à Z et des chiffres de 0 à 9

1.6 - Complémentaire d'un ensemble ou d'une suite de caractères

Pour préciser que certains caractères ne doivent pas apparaître dans le résultat, ils sont notés entre crochets, et le crochet ouvrant est suivi du caractère "^"⁵.

Exemple 1.6.1

`d[^a-z]`

recherche des caractères qui ne sont pas des lettres de l'alphabet minuscules

Exemple 1.6.2

`d'[^aeiou]`

recherche des chaînes de caractères commençant par "d apostrophe" non suivi des caractères "a", "e", "i", "o" ou "u"

1.7 - Répétitions

Trois caractères spécifient les répétitions. Ils portent sur l'entité qui les précède.

* : le caractère apparaît 0 ou plusieurs fois

+ : le caractère apparaît 1 ou plusieurs fois

? : le caractère apparaît 0 ou une fois

Exemple 1.7.1

`des?`

recherche des chaînes de caractères "de" ou "des"

⁵ Clavier azerty : appuyer sur la touche 'Alt Gr' et, en même temps sur la touche '9'.

Exemple 1.7.2

`[aeiou]+`

recherche des chaînes de caractères de une ou plusieurs voyelles

Exemple 1.7.3

`_que[^,.*]*`

recherche de
– un caractère espace
– la suite de caractères "que"
– des caractères qui ne sont ni une virgule, ni un espace, ni ou point.

Exercice 1.7.1 : Construire une expression rationnelle qui recherche les nombres.

Exercice 1.7.2 : Construire une expression rationnelle qui recherche les mots : marcher, marche, marches, marchent, chanter, chante, chantes, chantent, manger, mange, manges, mangent.

Exercice 1.7.3 : Construire une expression rationnelle qui recherche les mots : former, réformer, déformer, informer, forme, réforme, déforme, informe, formes, réformes, déformes, informes, forment, réforment, déforment, informent

1.8 - Caractère joker

Le caractère point " ." remplace n'importe quel caractère.

Exemple 1.8.1

`qu.`

recherche des chaînes de caractères commençant par "qu" suivies de n'importe quel autre caractère

1.9 - paragraphes

Le caractère de l'accent circonflexe " ^ ", lorsqu'il ne figure pas entre crochets, marque le début d'un paragraphe.

Le caractère " \$ " marque la fin d'un paragraphe.

Exemple 1.9.1

`^$`

Recherche un paragraphe qui ne contient pas de caractères

Exemple 1.9.2

`^.*!.*$`

recherche un paragraphe qui contient au moins un point d'exclamation

Exemple 1.9.3

`^[^!]*![^!]*$`

recherche un paragraphe qui contient un et un seul point d'exclamation

Exemple 1.9.4

`^(Le|La|Les)`

recherche des chaînes de caractères qui commencent un paragraphe par "Le", "La" ou "Les"

Exercice 1.9.1 : Rechercher le premier mot (écrit avec des lettres non accentuées) qui commence un paragraphe et qui est suivi d'une virgule ou d'un espace.

Exercice 1.9.2 : Rechercher le premier mot qui commence un paragraphe et qui est suivi d'une virgule ou d'un espace.

1.10. Caractère d'échappement

Pour rechercher un caractère ayant une interprétation spéciale (comme "*", "(", "[", "+", "|", ":", etc.), il faut les faire précéder du caractère d'échappement antislash "\"⁶ ou les inclure dans une suite de caractères.

Exemple 1.10.1

`\(.*\)`

recherche de

- la première parenthèse ouvrante
- n'importe quel nombre de fois un caractère quelconque (cela peut très bien être d'autres parenthèses)
- la dernière parenthèse fermante

Exemple 1.10.2

`\([\^()]*\)`

recherche de

- la parenthèse ouvrante
- n'importe quel nombre de fois un caractère quelconque mais qui n'est ni une parenthèse ouvrante "(", ni une parenthèse fermante ")"
- la parenthèse fermante

Exemple 1.10.3

`_[a-z]+\.`

recherche une espace suivie du dernier mot d'une phrase et d'un point

Exercice 1.10.1 : quelle est la différence entre les expressions rationnelles des exemples 1.10.1 et 1.10.2 ?

⁶ Clavier azerty : appuyer sur la touche 'Alt Gr' et, en même temps sur la touche '8'.

1.11. Méthodologie de construction d'une expression rationnelle

- 1 - Analyser la question en termes d'éléments successifs
- 2 - Convertir chaque élément constitutif avec la syntaxe des expressions rationnelles
- 3 - Concaténer chacun des éléments

Exemple : trouver les étiquettes lexicales. Une étiquette lexicale est terminée par un point et est toujours située après la phonologie dans des entrées de dictionnaire du type :

balastu [bàlàstú] kkye2. Balas dio. Daudu sèram kare kasuwulan kiladènniya, bune ti-a banju ngimba-a balaskera.[Fa.: compte (faire le), bilan (faire le)]

fula feye [fúlà fêyè] cuk. Kiam kèndèrmu jagandène awo kèndawua samenjun kiaccaidè. Fularo bër-a lawasar-a kèljagane dejaiya, jairo buwucciro walji. M.: kèndáwù fêyè.[Fa.: beurre]

gajawa [gàjàwá] cu. Tulo ingia, dangoadè. Motaa curo gajawayero kèrgayiyendeya, luwo tiyero.[Fa.: boue]

1 – Analyse

Il faut chercher

- un crochet fermant
- un ou plusieurs espaces
- un ou plusieurs caractères différents du point
- un point

2 – Conversion de chaque élément avec la syntaxe des expressions rationnelles

- \] un crochet fermant
- _+ un ou plusieurs espaces
- [^\.]+ un ou plusieurs caractères différents du point
- \. un point

3 – Concaténation des éléments pour former l'expression rationnelle

\] _+ [^\.]+ \.

2 - Expressions rationnelles : modification

Une expression rationnelle permet de localiser une chaîne de caractères et de la modifier.

Notations

L'expression rationnelle est surlignée en **vert pâle** et la chaîne de substitution est surlignée en **orange** pour plus de clarté. Il faut faire attention à la lecture de l'expression quand elle comporte des espaces.

Exercices pratiques

Les expressions rationnelles seront utilisées avec l'éditeur de texte Open Office⁷.

Utiliser la fonction "Rechercher & Remplacer" de l'éditeur de texte :

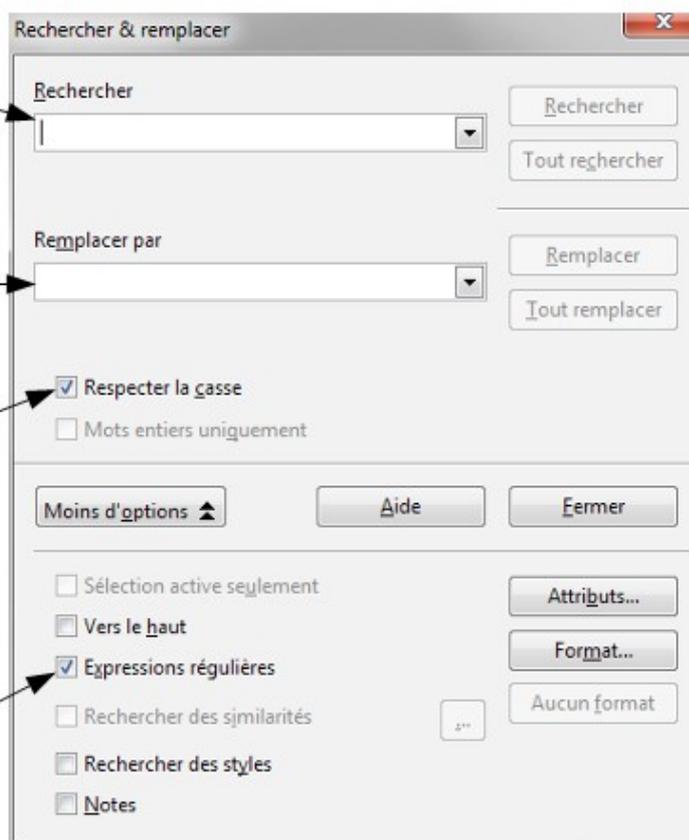
Édition > Rechercher & Remplacer⁸

Placer la chaîne de caractère à localiser dans la case « Rechercher »

Placer l'expression de remplacement dans la case « Remplacer par »

Cocher la case "Respecter la casse" afin de distinguer caractères minuscules et majuscules

Cocher la case "Expressions rationnelles" dans les options.



⁷ Téléchargeable gratuitement : <http://fr.openoffice.org/>

⁸ Pour ouvrir la fenêtre avec le raccourci, appuyer sur les touches 'Ctrl' et 'F' en même temps.

Taper les exemples dans la case "Rechercher" et les chaînes de remplacement dans "Remplacer par" et regarder les résultats obtenus en cliquant sur le bouton "Remplacer" ou "Tout remplacer".

2.1 - Substitution totale

La chaîne correspondant à l'expression rationnelle est remplacée par la chaîne de substitution.

Exemple 2.1

`([aeiou])ss`

`Aaaaass`

2.2 - Substitution partielle

Une sous-chaîne entre parenthèses peut être récupérée via une variable nommée par le signe dollar "\$", suivi d'un numéro.

Chaque ouverture de parenthèse est numérotée en commençant par le chiffre 1.

La sous-chaîne reconnue dans les parenthèses dont la parenthèse ouvrante est numérotée 'j' est nommée "\$j".

Exemple 2.2.1

`([aeiou])ss`

`$1$1ss`

Exemple 2.2.2

`([a-z]*)_1([a-z]*)_2([a-z]*)_3`

`$1-$3`

Exemple 2.2.3

`([aeiou])ss([aeiou])`

`$2X$1`

Exemple 2.2.4

`_1([a-zA-Z])_2([a-z])_3([a-z])`

`_$3$2$1`

2.3 - Méthodologie de construction d'une expression rationnelle avec substitution

- 1 – Analyser la question en termes d'éléments successifs
- 2 – Convertir chaque élément constitutif avec la syntaxe des expressions rationnelles
- 3 – Repérer les éléments à conserver (mettre dans les variables) et les encadrer de parenthèses
- 4 – Concaténer chacun des éléments
- 5 – Écriture de l'expression de remplacement

Exemple : Encadrer les étiquettes lexicales par les balises <étiquette> et </étiquette>. Une étiquette lexicale est terminée par un point et est toujours située après la phonologie dans des entrées de dictionnaire du type :

balastu [bàlâstú] kkye2. Balas dio. Daudu sèram kare kasuwulan kiladènniya, bune ti-a banju ngimba-a balaskera.[Fa.: compte (faire le), bilan (faire le)]
fula feye [fúlà fêyè] cuk. Kiam kèndèrmu jagandène awo kèndawua samenjun kiaccaidè. Fularo bër-a lawasar-a kèljagane dejaiya, jauru buwucciro walji. M.: kèndáwù fêyè.[Fa.: beurre]
gajawa [gàjàwá] cu. Tulo ingia, dangoadè. Motaa curo gajawayero kèrgayiyendeya, luwo tiyero.[Fa.: boue]

1 – Analyse :

Il faut chercher

un crochet fermant
un ou plusieurs espaces
un ou plusieurs caractères différents du point : début de l'étiquette
un point : termine l'étiquette

2 – Conversion de chaque élément avec la syntaxe des expressions rationnelles

\] un crochet fermant
_+ un ou plusieurs espaces
[^ .]+ un ou plusieurs caractères différents du point : début de l'étiquette
\. un point : termine l'étiquette

3 – Repérer les éléments à conserver (mettre dans les variables) et les encadrer de parenthèses

Ici, il faut récupérer ce qui constitue l'étiquette :

\] un crochet fermant
_+ un ou plusieurs espaces
([^ .]+) un ou plusieurs caractères différents du point : début de l'étiquette
(\.) un point : termine l'étiquette

4 – Concaténation des éléments pour former l'expression rationnelle

\]_+([^ .]+)(\.)

5 – Écriture de l'expression de remplacement

Le crochet fermant précédant l'étiquette doit être conservé, ainsi qu'un espace.

L'étiquette contenue dans les variables est encadrée par des balises.

]_<etiquette>\$1\$2</etiquette>

En appliquant cette expression rationnelle et la chaîne de remplacement les entrées précédentes deviennent :

balastu [bàlâstú] <etiquette>kkye2.</etiquette> Balas dio. Daudu sèram kare kasuwulan kiladènniya, bune ti-a banju ngimba-a balaskera.[Fa.: compte (faire le), bilan (faire le)]
fula feye [fúlà fêyè] <etiquette>cuk.</etiquette> Kiam kèndèrmu jagandène awo kèndawua samenjun kiaccaidè. Fularo bër-a lawasar-a kèljagane dejaiya, jauru buwucciro walji. M.: kèndáwù fêyè.[Fa.: beurre]
gajawa [gàjàwá] <etiquette>cu.</etiquette> Tulo ingia, dangoadè. Motaa curo gajawayero kèrgayiyendeya, luwo tiyero.[Fa.: boue]

Remarque :

L'expression rationnelle suivante est équivalente (l'étiquette est regroupée avec son point dans une même variable) :

```
\]_+([\^.\.]+\.)
```

```
]_<etiquette>$1</etiquette>
```

3 – Correction des exercices

Exercice 1.4 : Construire une expression rationnelle qui recherche les mots d'une lettre suivis d'une voyelle élidée comme "d'", "m'", "t'", etc.

`[dmtl](qu)'`

Exercice 1.7.1 : Construire une expression rationnelle qui recherche les nombres.

`[0-9]+'`

Exercice 1.7.2 : Construire une expression rationnelle qui recherche les mots : marcher, marche, marches, marchent, chanter, chante, chantes, chantent, manger, mange, manges, mangent.

`(marche|chante|mange)(r|s|nt)?`

Exercice 1.7.3 : Construire une expression rationnelle qui recherche les mots : former, réformer, déformer, informer, forme, réforme, déforme, informe, formes, réformes, déformes, informes, forment, réforment, déforment, informent

`(ré|dé|in)?forme(r|s|nt)?`

Exercice 1.9.1 : Rechercher le premier mot (écrit avec des lettres non accentuées) qui commence un paragraphe et qui est suivi d'une virgule ou d'un espace.

`^[^àâéèêîïù](,|_)`

Exercice 1.9.2 : Rechercher le premier mot qui commence un paragraphe et qui est suivi d'une virgule ou d'un espace.

`^[^*] (,|_)`

Exercice 1.10.1 : Quelle est la différence entre les expressions rationnelles des exemples 1.10.1 `\(.*\)` et 1.10.2 `\([^(]*)*\)` ?

La première expression recherche la toute première parenthèse ouvrante d'une ligne et la dernière parenthèse fermante de cette ligne. Le résultat est tout ce qui se trouve à l'intérieur de cette sélection, même s'il existe d'autres parenthèses avant. L'outil de recherche les interprète comme étant des caractères quelconques répétés plusieurs fois à cause de la sous-chaîne ".*"

La seconde expression en revanche spécifie que ces caractères quelconques ne doivent pas être d'autres parenthèses (ouvrante ou fermante), donc la recherche s'arrête à la première parenthèse fermante rencontrée, qui correspond au dernier caractère de l'expression.