

Utiliser les expressions régulières

Estelle Debouy

1 Définition

On appelle « expressions régulières » un certain nombre de signes conventionnels, utilisables par exemple avec la fonction « Rechercher/remplacer » de LibreOffice, qui permettent de retrouver des séquences complexes de caractères dans un document.

Pour que le logiciel prête aux signes en question leur valeur « d’expressions régulières », et non leur sens habituel (\wedge = accent circonflexe, \$ = dollar, etc.), il ne faut pas omettre de cocher la case « expressions régulières » dans la boîte « Rechercher/Remplacer » de LibreOffice, sous le bouton « Plus d’options ».

Voici quelques expressions régulières :

- \wedge = début de paragraphe. Exemple : la recherche « \wedge Roméo » sélectionnera le mot « Roméo » seulement s’il commence un paragraphe.
- \$ = fin de paragraphe. Exemple : la recherche « Roméo\$ » sélectionnera le mot « Roméo » seulement s’il termine un paragraphe.
- \wedge \$ = deux marques de paragraphe consécutives.
- $\backslash n$ = une marque de paragraphe
- $\backslash t$ = une tabulation
- . = un caractère quelconque

2 Exemples d’application

Application n° 1 Vous avez peut-être constaté qu’après avoir récupéré une œuvre en prose depuis une bibliothèque numérique (Gallica par exemple), il faut très souvent supprimer les retours à la ligne intempestifs qu’on appelle *hard wrap*. Exemple :

THE-GOLDEN-BIRD¶
 ¶
 A-certain-king-had-a-beautiful-garden,-and-in-the-garden-stood-a-tree¶
 which-bore-golden-apples.-These-apples-were-always-counted,-and-about¶
 the-time-when-they-began-to-grow-ripe-it-was-found-that-every-night-one¶
 of-them-was-gone.-The-king-became-very-angry-at-this,-and-ordered-the¶
 gardener-to-keep-watch-all-night-under-the-tree.-The-gardener-set-his¶
 eldest-son-to-watch;-but-about-twelve-o'clock-he-fell-asleep,-and-in¶
 the-morning-another-of-the-apples-was-missing.-Then-the-second-son-was¶
 ordered-to-watch;-and-at-midnight-he-too-fell-asleep,-and-in-the-morning¶
 another-apple-was-gone.-Then-the-third-son-offered-to-keep-watch;-but¶
 the-gardener-at-first-would-not-let-him,-for-fear-some-harm-should-come¶
 to-him;-however,-at-last-he-consented,-and-the-young-man-laid-himself¶
 under-the-tree-to-watch.-As-the-clock-struck-twelve-he-heard-a-rustling¶
 noise-in-the-air,-and-a-bird-came-flying-that-was-of-pure-gold;-and-as¶
 it-was-snapping-at-one-of-the-apples-with-its-beak,-the-gardener's-son¶
 jumped-up-and-shot-an-arrow-at-it.-But-the-arrow-did-the-bird-no-harm;-¶
 only-it-dropped-a-golden-feather-from-its-tail,-and-then-flew-away.¶
 The-golden-feather-was-brought-to-the-king-in-the-morning,-and-all-the¶
 council-was-called-together.-Everyone-agreed-that-it-was-worth-more-than¶
 all-the-wealth-of-the-kingdom;-but-the-king-said,-'One-feather-is-of-no¶
 use-to-me,-I-must-have-the-whole-bird.¶
 ¶

Pour supprimer ces retours à la ligne, le plus simple est d'utiliser un éditeur comme Notepad++ : allez dans le menu TextFX, commande TextFXEdit, sous-commande Unwrap Text. Les retours à la ligne simples sont convertis en fin de ligne (*soft wrap*) tandis que les doubles retours subsistent.

Mais il est possible de rester dans LibreOffice et d'utiliser les expressions régulières. Notez d'abord qu'il ne faudra pas supprimer tous les retours à la ligne. Là où il y a plusieurs retours à la ligne (symbolisés par au moins deux pieds de mouche ¶), il faudra conserver un retour à la ligne :

1. vous devez donc commencer par rechercher la séquence ~\$ (expression qui désigne les sauts de paragraphes) et la remplacer par le caractère § par exemple, de façon à les marquer ;
2. puis recherchez la séquence \$ (expression qui désigne les retours à la ligne) et remplacez-la par une espace : tous les retours à la ligne du texte sont alors remplacés. Vous obtenez alors un texte « au kilomètre » ;
3. il est maintenant temps de réintroduire les sauts de paragraphes : recherchez le caractère § et remplacez-le par la séquence \n (qui introduit un retour à la ligne là où l'on voulait qu'il en restât).

Application n° 2 Les expressions régulières sont aussi très utiles pour appliquer un formatage particulier à un terme donné, seulement quand il se trouve dans une certaine position. Prenons l'exemple d'une pièce de théâtre téléchargée depuis une bibliothèque numérique : comment retrouver les occurrences des noms de personnages servant à indiquer les prises de parole (distinctes des occurrences où le nom est effectivement prononcé par un

personnage) afin de les définir en tant que paragraphes, ce qui permettra de leur appliquer des petites majuscules ?

1. pour chaque personnage, recherchez les passages où il se trouve en début de ligne et ajoutez une marque de paragraphe après : recherchez « ^Père Ubu » et remplacez par « Père Ubu\n » ;
2. vous pouvez alors rechercher toutes les occurrences qui sont maintenant isolées dans un paragraphe afin de leur appliquer des petites majuscules : recherchez ^Père Ubu\$ (et pensez à choisir Tout rechercher).

Pour une liste plus complète des expressions régulières, voir https://help.libreoffice.org/Common/List_of_Regular_Expressions/fr.