

[↑ \(/bin/view/Main/\) / Documentation Frantext \(/bin/view/Main/\) / Manuel d'utilisation \(/bin/view/Main/Manuel%20d%27utilisation/\)](#)  
[/ Expressions \(/bin/view/Main/Manuel%20d%27utilisation/Expressions/\)](#)  
[/ Les expressions régulières \(/bin/view/Main/Manuel%20d%27utilisation/Expressions/Les%20expressions%20re%CC%81gulie%CC%80res/\)](#)

# Les expressions régulières

Le moteur de recherche de Frantext permet l'utilisation d'expressions régulières – appelées également expressions rationnelles – pour rechercher des suites de caractères selon des motifs et des formules logiques. Les expressions régulières peuvent être utilisées dans les recherche avancées, de fréquences, de co-occurrences et de voisinages, ainsi que dans les listes de mots et les grammaires.

## DANS CETTE PAGE

- Opérateurs
- Quantificateurs
- Classes de caractères

## Opérateurs

Les opérateurs permettent de spécifier le type de caractères à rechercher.

Opérateur	Description	Exemple de recherche avancée	Exemple de résultats séparés par des virgules
()	<b>Groupe de caractères</b> d'une expression	"(nuit)"	nuit
	<b>Choix</b> entre plusieurs alternatives	"jour nuit"	jour, nuit
.	<b>N'importe quel caractère</b>	"n.it"	nuit, nait...
\	<b>Interpréter littéralement un opérateur</b>	"\"."	.
[]	<b>Un des caractères entre crochets</b>	"[bp]eau"	beau, peau
[^]	<b>Tout caractère hormis ceux entre crochets</b>	"n[^u]is"	nais, nois
[a-z]	<b>Un intervalle composé de caractères alphabétiques de a à z, en minuscules et sans diacritiques</b>	"[a-z]ait"	fait, sait, lait...
[0-9]	<b>Un intervalle composé de chiffres de 0 à 9</b>	"[0-9]"	0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Pour aller plus loin, consultez la documentation dédiée à l'utilisation des caractères spéciaux ([/bin/view/Main/Manuel%20d%27utilisation/Expressions/Les%20caracte%CC%80res%20spe%CC%81ciaux/](#)) dans Frantext.

## Quantificateurs

Les quantificateurs permettent de spécifier le nombre de caractères à rechercher.

Quantificateur	Description	Exemple de recherche avancée	Exemple de résultats séparés par des virgules
?	<b>Zéro ou une fois</b> le caractère ou groupe qui précède	"nu(it)?"	nuit, nu
*	<b>Zéro ou plusieurs fois</b> le caractère ou groupe qui précède	"cré*e"	cre, crée, créée
+	<b>Une ou plusieurs fois</b> le caractère ou groupe qui précède	"cré+e"	créée, créée
{n}	<b>Exactement n occurrences</b> de l'expression précédant les accolades. La valeur de n est limitée à 32.	"(ha){2}"	haha
{n,n}	<b>Exactement n occurrences</b> de l'expression précédant les accolades. La valeur de n est limitée à 32.	"(ha){2,2}"	haha
{n,m}	<b>Entre n et m occurrences</b> de l'expression précédant les accolades. Les valeurs de n et m sont limitées à 32.	"(ha){2,3}"	haha, hahaha
{n,}	<b>Au moins n occurrences</b> de l'expression précédant les accolades. La valeur de n est limitée à 32.	"(ha){2,}"	haha, hahaha, hahahaha...

## Classes de caractères

Les classes de caractères correspondent à des caractères ou des intervalles prédéfinis.

Classe	Description	Exemple de recherche avancée	Exemple de résultats séparés par des virgules
<code>\d</code>	<b>Un chiffre</b> , correspond à <code>[0-9]</code>	<code>"\d"</code>	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
<code>\D</code>	<b>Tout caractère hormis un chiffre</b> , correspond à <code>[^0-9]</code>	<code>"\D"</code>	., à, -, a, ?, ;, y, ;, «...
<code>\x</code>	<b>Nombre hexadécimal sur 2 chiffres</b>	<code>"\x40"</code>	@
<code>\u</code>	<b>Nombre hexadécimal sur 4 chiffres</b> , correspond aux caractères unicodes	<code>"\u0040"</code>	@
<code>\S</code>	<b>Tout caractère hormis un espace</b> , correspond à <code>[\t\n\f\r ]</code>	<code>"\S"</code>	., à, -, a, ?, ;, y, ;, «...
<code>\w</code>	<b>Un caractère alphanumérique</b> sans diacritiques, correspond à <code>[0-9A-Za-z_]</code>	<code>"\w"</code>	a, y, 2, 1, s, h, 4, 3...
<code>\W</code>	<b>Tout caractère hormis un caractère alphanumérique</b> , correspond à <code>[^0-9A-Za-z_]</code>	<code>"\W"</code>	., -, ?, ;, ;, «...
<code>:alnum:</code>	<b>Un caractère alphanumérique</b> sans diacritiques, correspond à <code>[0-9A-Za-z]</code>	<code>"[:alnum:]"</code>	a, y, 2, 1, s, h...
<code>:alpha:</code>	<b>Un caractère alphabétique</b> sans diacritiques, correspond à <code>[A-Za-z]</code>	<code>"[:alpha:]"</code>	a, y, s, h, i, t...
<code>:digit:</code>	<b>Un caractère numérique</b> , correspond à <code>[0-9]</code>	<code>"[:digit:]"</code>	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
<code>:lower:</code>	<b>Un caractère alphabétique en minuscules</b> sans diacritiques, correspond à <code>[a-z]</code>	<code>"[:lower:]"</code>	a, y, s, h, i, t...
<code>:upper:</code>	<b>Un caractère alphabétique en majuscules</b> sans diacritiques, correspond à <code>[A-Z]</code>	<code>"[:upper:]"</code>	m, a, i, p, b, c...
<code>:word:</code>	<b>Un caractère alphanumérique ou un tiret bas</b> , correspond à <code>[0-9A-Za-z_]</code>	<code>"[:word:]"</code>	a, y, 2, 1, s, h, 4, 3...
<code>:xdigit:</code>	<b>Un nombre hexadécimal</b> , correspond à <code>[0-9A-Fa-f]</code>	<code>"[:xdigit:]"</code>	a, 2, 1, 3, 4, 6, 5...

Le moteur de recherche de Frantext (baptisé *Allegro*) est également capable d'exploiter des classes de caractères correspondant à des éléments invisibles, bien que ceux-ci ne puissent pas être exploités dans l'interface en ligne de Frantext du fait de la segmentation des mots.

Classe	Description
<code>\t</code>	<b>Tabulation horizontale</b>
<code>\r</code>	<b>Retour chariot</b>
<code>\n</code>	<b>Saut de ligne</b>
<code>\f</code>	<b>Saut de page</b>
<code>\s</code>	<b>Un caractère d'espace</b> , correspond à <code>[\t\n\f\r ]</code>
<code>:blank:</code>	<b>Un caractère de tabulation</b> , correspond à <code>[\t ]</code>
<code>:space:</code>	<b>Un caractère d'espacement</b> , correspond à <code>[\t\n\f\r ]</code>

Tags : manuel (</bin/view/Main/Tags?do=viewTag&tag=manuel>) Allegro (</bin/view/Main/Tags?do=viewTag&tag=Allegro>) v1.2.21 (</bin/view/Main/Tags?do=viewTag&tag=v1.2.21>)

Modifié par Gilles Toubiana (</bin/view/XWiki/gtoubiana>) le 2019/07/05 09:17