

Expressions FRANTEXT 19.2

Résumé des expressions de la recherche avancée

Expressions CQL (Corpus Query Language)

Forme graphique du mot \$	[word="\$ "] ou "\$ "
Flexions du lemme \$	[lemma="\$ "]
Catégorie grammaticale \$	[pos="\$ "]
Joker, n'importe quel mot	[]
\$1 suivi de \$2	[word="\$1 "] [word="\$2 "]
\$1 et \$2	[word="\$1 " & pos="\$2 "]
\$1 ou \$2	[word="\$1 " word="\$2 "]
\$1 dont la catégorie grammaticale n'est pas \$2	[lemma="\$1 " & pos!="\$2 "]
\$1 entre les balises <\$2> et </\$2>	<\$2> [lemma="\$1 "] </\$2>
Recherche sensible à la casse des caractères	[word="\$ "]
Recherche insensible à la casse des caractères	[word="\$ "%c]
Recherche insensible aux signes diacritiques	[word="\$ "%d]
Recherche insensible aux ligatures	[word="\$ "%g]
Recherche littérale, insensible aux expressions régulières	[word="\$ "%l]
Cumul de modificateurs	[word="\$ "%cdgl]

Expressions de séquence

Liste dont l'identifiant est \$	&liste("\$ ")
Règle \$2 de la grammaire \$1	&règle("\$1 :: \$2 ")
Flexions du lemme \$1 issues du lexique \$2 (médiéval, préclassique ou moderne). Les catégories grammaticales \$3 et \$4 sont optionnelles	&lexique("\$1 " , "\$2 " , "\$3 ") ou [word=&lexique("\$1 " , "\$2 " , "\$3 ") & pos="\$4 "]
Recherche de multiples unités lexicales \$	&segmenter("\$ ")
Recherche simple de la phrase \$	&développer("\$ ")

Catégories grammaticales

Adjectif	ADJ
Adverbe	ADV
Conjonction de coordination	CC
Conjonction de subordination	CS
Clitique objet	CLO
Clitique sujet	CLS
Déterminant	DET
Mot étranger	ET
Interjection	I
Nom commun	NC
Nom propre	NP
Préposition + déterminant	P\+D
Ponctuation	PONCT
Pronom	PRO
Pronom relatif	PROREL
Pronom interrogatif	PROWH
Préposition	P
Verbe conjugué	V
Verbe à l'infinitif	VINF
Verbe participe passé	VPP
Verbe participe présent	VPR
Mot non traité	X
Adverbe + clitique objet	ADV\CLO
Adverbe + déterminant	ADV\DET
Clitique sujet + clitique objet	CLS\CLO
Clitique sujet + verbe	CLS\+V
Conjonction de subordination + clitique	CS\CLO
Conj. de subordination + clitique sujet	CS\+CLS
Verbe + clitique sujet	V\+CLS

Expressions régulières

Groupe (\$)	
Alternative (\$1 ou \$2)	\$1 \$2
N'importe quel caractère	.
Interpréter littéralement un opérateur	\
Un des caractères entre crochets	[\$]
Tout caractère hormis ceux entre crochets	[^ \$]
Intervalle alphabétique	[a-z]
Intervalle numérique	[0-9]
0 ou 1 fois	\$?
0 fois ou plus	\$ *
1 fois ou plus	\$ +
Exactement n fois	\$ {n}
Entre n et m fois	\$ {n,m}
Au moins n fois	\$ {n,}
Un chiffre	\d
Tout sauf un chiffre	\D
Nombre hexadécimal sur 2 chiffres	\x
Nombre hexadécimal sur 4 chiffres	\u
Tout sauf un espace	\S
Un caractère alphanumérique sans diacritiques	\w
Tout sauf un caractère alphanumérique	\W

Un caractère alphanumérique (1)	[[:alnum:]]
Un caractère alphabétique (1)	[[:alpha:]]
Un caractère numérique	[[:digit:]]
Un caractère alphabétique en minuscules (1)	[[:lower:]]
Un caractère alphabétique en majuscules (1)	[[:upper:]]
Un caractère alphanumérique ou un tiret bas	[[:word:]]
Un nombre hexadécimal	[[:xdigit:]]

(1) Sans les caractères diacritiques