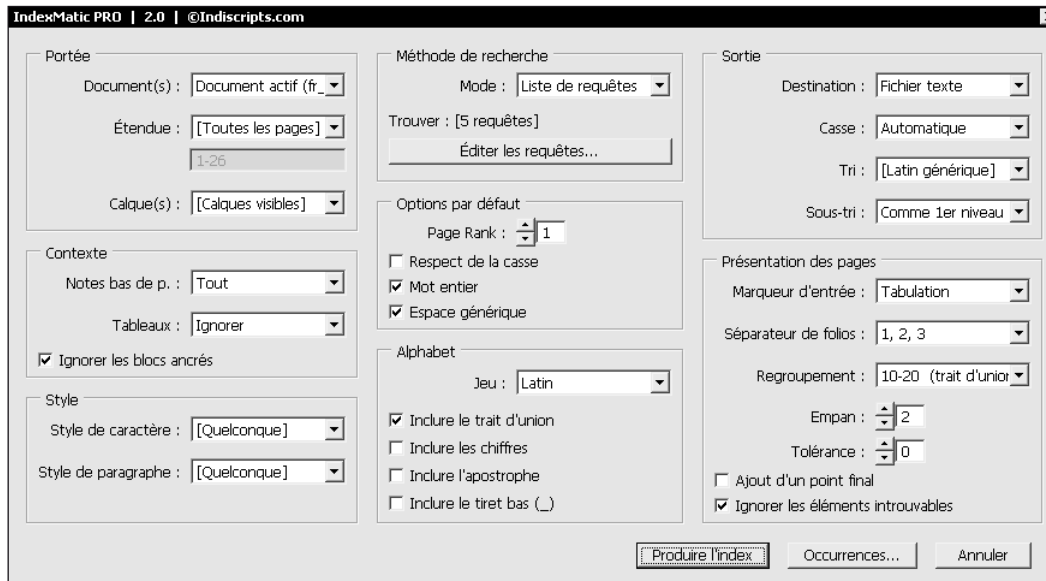


IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Prise en main



Boîte de dialogue principale d'IndexMatic² (v. 2.025, sous Windows).

1 INSTALLATION ET LANCEMENT D'INDEXMATIC²

1. Téléchargez IndexMatic² depuis indiscripts.com.
[Version PRO] Téléchargez le fichier via votre lien privé et sauvegardez-le sur disque.
[Version TRY] Clic droit sur [IndexMaticTry.zip](#), puis «Enregistrer la cible sous...»
2. Après extraction de [IndexMaticPro.jsx](#) (ou [...Try.jsx](#)) depuis le fichier ZIP, placez le fichier dans votre dossier de scripts InDesign (habituellement: **Scripts Panel**).
3. Démarrez InDesign, ouvrez un document. Depuis la palette Scripts — **Window > Utilities > Scripts** (CS5) ou **Window > Automation > Scripts** (CS3, CS4) — double-cliquez sur [IndexMatic...jsx](#).

- IndexMatic² supporte **InDesign CS3, CS4 ou CS5+** (recommandé) et fonctionne sur les plateformes **Mac** et **Windows**.
- La **version TRY** est soumise à quelques limitations: vous ne pouvez pas changer le format de destination (verrouillé sur **Fichier texte**), l'index final n'accepte que **50 termes**, la fonction **Occurrences...** est masquée, enfin le script ne mémorise pas les **derniers paramètres utilisés**.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Prise en main

4 STRUCTURE

Micromètre, le leur fait en réalité une solide dose de bon sens pour mener à bien une étude technique. Le dimensionnement d'un bâtiment en boîtes de paille s'échappe par là comme d'habitude. En tant que concepteur d'un projet de construction, vous devez trouver l'équilibre entre physique, science des matériaux et réalités de l'industrie locale du bâtiment afin d'arriver à des conclusions raisonnables pour votre projet. C'est tout l'art de l'ingénieur.

Cela dit, comme dans de nombreux ouvrages techniques ou codes du bâtiment, nous vous recommandons, à titre préventif mais encore proviseur, des valeurs de charge minimales dans les tableaux de la section 5. Ce document est en effet encore à l'état de projet. Il doit évoluer et passer par de nombreuses révisions avant de prétendre faire partie du Code national de la construction; ce n'est pas un document officiel mais il apporte à l'éleveur. Les valeurs qui y sont données, doivent également être utilisées avec beaucoup de bon sens.

Après toutes ces recommandations et réserves, il convient de souligner que nous avons beaucoup appris en quelques années sur la façon de concevoir et de dimensionner des bâtiments en boîtes de paille, et que nous pouvons désormais le faire avec un certain degré de confiance, partout dans le monde. Il est certain que nous avons encore beaucoup à apprendre et que de nombreux essais sont encore à réaliser, notamment de fait de l'évaluation technique que connaît actuellement la construction en boîtes de paille. Mais aujourd'hui, on ne se lance plus à l'aveuglette; on a simplement répondu aux grandes interrogations, reste maintenant à résoudre la multitude de petites questions qui se posent encore.

Le témoignage essentiel, et non encore cité, est peut-être celui fourni par la nature. Il existe aujourd'hui des structures portantes en boîtes de paille partout dans le monde, certaines ont plus de cent ans; elles nous indiquent à partir d'une construction de boîtes, d'adultes et de matériaux de plus variés, mais nous n'avons encore jamais entendu parler d'aucun effondrement. Cette observation est destinée à vous signaler sur le choix d'un coefficient de sécurité, pour votre projet. Les murs en boîtes de paille enduits présentent une solidité et une réserve de résistance qui n'ont été que pressenties par les essais conduits à ce jour, mais qui sont attestées par les maisons continues de Nebraska. Soyez prudent — oui, ce n'est pas toujours le fait ou le don de votre être! — mais prenez votre décision en gardant un œil sur les résultats d'essai et l'attente sur les inévitables expériences de terrain, signez! Nous sommes reconnaissants.

5

A

B

C

à répéter les essais ou à régler le plan, compléments sont offerts de décaler, à l'échelle ou...

IndexMatic PRO | Éditeur de requêtes

Éditer les requêtes :

- calcaire
- carbone
- chanvre
- /absorb(b|p)\w+ / => absorption
- /enduit\w*/5 => enduit
- /bottes*/5 => boîte de paille
- cellulose
- chaleur
- /flammes|feu|incendie? / => feu
- /Afrique|Amérique|France|Canada / => Géographie > \$0
- /architectes? / => architecte
- /argiles? / => argile
- bactéries
- /Californi(e|a) / => Géographie > Californie
- /drains(nt|ge) / => drainage
- bardage
- /briques? / => brique

Pressez Ctrl+Enter pour ajouter une ligne

OK Annuler

1	absorption	47, 54, 58-59, 66, 78-80, 89, 13
2	acoustique	22, 90, 218, 235-241, 313, 319
3	architecte	13-14, 16, 18, 20-22, 34, 67, 11
4	argile	31-32, 45, 48-51, 55-59, 61, 81
5	bactéries	158, 161, 166, 169-171
6	bardage	59, 182, 184-185, 189, 191, 193
7	boîte de paille	24, 27-30, 33, 38, 40-41, 47, 61
8	brique	16, 28, 32, 38, 45, 78, 86, 113
9	calcaire	51-52, 54, 81
10	carbone	16, 26-27, 30, 47, 51, 54, 78, 1
11	cellulose	26, 48, 66, 88, 90, 158, 174, 22
12	chaleur	19, 25, 40, 43, 46-47, 54, 59, 1
13	chanvre	28, 38, 48, 66-70, 73, 75-77, 13
14	charbon	27, 233
15	charge	9, 29, 32-33, 36, 38, 42, 46-49,
16	ciment	16-17, 21, 32, 36, 38-39, 45, 48
17	combustion	171, 209, 214-217
18	densité	32-33, 36-37, 47, 67, 85, 99-106
19	dioxyde	51, 54, 156, 161-163, 171
20	drainage	181-182, 192-194, 198, 254, 311,
21	enduit	9, 45-50, 52, 54-55, 58-61, 63-6
22	énergie	2, 14-15, 22-24, 47, 52, 54, 57,
23	étanchéité	21, 58, 63, 69, 77-78, 80, 90, 9
24	façade	60, 186, 189-191, 257, 319
25	feu	19, 21, 26, 37, 41, 46-47, 63, 8
26	fibre	38-39, 67, 88, 103, 106, 139, 15
27	finition	21-22, 36, 41, 45, 55, 60, 63, 6
28	fissure	9, 49, 65-68, 77-79, 81, 91, 105
29	fluage	101, 113-117, 309, 317
30	GÉOGRAPHIE	
31	Afrique	13, 268-269, 314
32	Amérique	33, 67, 200, 244, 246, 270
33	Californie	10, 12-16, 18, 22, 28, 30, 39, 4
34	Canada	16, 40, 42, 194, 274, 285, 304,

- IndexMatic² n'interagit pas avec les fonctionnalités d'index disponibles nativement dans InDesign. Le script ne modifie jamais les documents existants (il se contente de les inspecter).
- IndexMatic² offrira généralement des performances très supérieures à celles d'autres outils de même nature, car il ne sollicite pas le moteur GREP d'InDesign.

2 APERÇU RAPIDE

- IndexMatic² vous permet de produire un **index indépendant** à partir d'un document ou livre InDesign **A**. Le script est spécialement adapté à la production d'index thématiques, tables d'auteurs, glossaires, etc. Il offre divers moyens d'explorer en profondeur les documents cibles, de détecter les occurrences de mots ou expressions recherchés, puis de dresser une table de correspondance avec les numéros de page (folios).
- IndexMatic² s'appuie sur un **moteur de requêtes** sophistiqué **B**, permettant un ciblage très fin par expressions régulières (regex), filtrage selon les styles, 'page rank', réécriture automatique des termes...
- La version PRO propose plusieurs formats de sortie: **fichier texte** **C**, **XML**, **extrait** InDesign (CS4/CS5).

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Prise en main

A Portée
Document(s) : Document actif (Te)
Étendue : [Quel(s) document(s) désirez-vous] [Toutes les pages]
1-320
Calque(s) : [Calques visibles]

B Méthode de recherche
Mode : Automatique
Longueur mini : 6
Longueur maxi : 10
Options par défaut
Page Rank : 1

C Nombre maximum de caractères par

D Options par défaut
Page Rank : 3
Le 'Page Rank', de 1 à 9, représente le nombre minimal d'occurrences par page d'une expression donnée pour que celle-ci soit prise en compte au numéro de page correspondant. Une requête particulière peut toujours supplanter le 'Page Rank' par défaut grâce à la syntaxe 'xxx/1', 'xxx/2', etc.

E

170 fissures	81, 134, 150, 216-217
171 flammes	211, 216
172 flèche	121
173 flexion	122-123, 138-139
174 fluage	116-117
175 fonction	57
176 fonctions	46
177 fondations	34, 90, 256
178 formule	65
179 France	9, 264-265
180 fuites	184, 198, 204
181 galvanise	140
182 garantir	226
183 gazeuse	156, 173
184 gestion	311
185 glissement	144-145
186 goutte	59
187 grande	76
188 grappaille	304
189 grosse	25
190 grosses	33
191 groupe	115, 237
192 habitat	2, 14
193 habitation	272
194 hauteur	129, 142, 150-151
195 humide	159-160, 171, 197
196 hydrogène	156
197 hydrophile	176
198 impacts	119
199 important	184
200 incendie	208, 212-213, 217
201 incendies	212
202 indice	236-237, 273
203 initial	116
204 insectes	243-244, 247
205 interne	175

3

INDEXER VOTRE PREMIER DOCUMENT EN 15 SECONDES

La façon la plus simple de tester IndexMatic² est de choisir le mode de recherche **Automatique**. Ce dernier masque la complexité du système et permet de produire très rapidement un index sans fournir aucun mot-clé.

1. Ouvrez un document InDesign et exécutez le script. Assurez-vous que « **Document actif** » est sélectionné dans la liste de Document(s) de la rubrique **Portée** **A**.
2. Dans la zone **Méthode de recherche**, sélectionnez le mode « **Automatique** » **B** et ajustez les longueurs **Mini** et **Maxi** **C** des mots à capturer.
3. Dans **Options par défaut**, fixez le **Page Rank** à 3 ou 4 **D**. Cliquez enfin sur le bouton **Produire l'index** **E**.

- En coulisses, le mode de recherche **Automatique** configure et envoie au module principal une série d'**expressions régulières** (regex) →15.
- Pour afficher uniquement les termes (sans indication de folios), cliquez sur le bouton **Occurrences...** →29 (Cette fonctionnalité est exclusive à la version PRO.)

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Portée et contexte

Document actif (TestMEP.indd)
Tous les documents visibles (3)
fr_IndexMatic-Manuel.indd
Bovary.indd

[Toutes les pages]
Pages...
Section:1
Section:2
Section:3
Section:4
Section:5
Section:6
Section:7
Section:8
Section:9
Section:10
Section:11
Section:12
Section:13
Section:14

[Tous]
[Calques visibles]
Texte
Legendage
Icône

4 DÉFINIR LA « PORTÉE »

La rubrique **Portée** **A** permet de circonscrire les documents à explorer et le champ d'indexation.

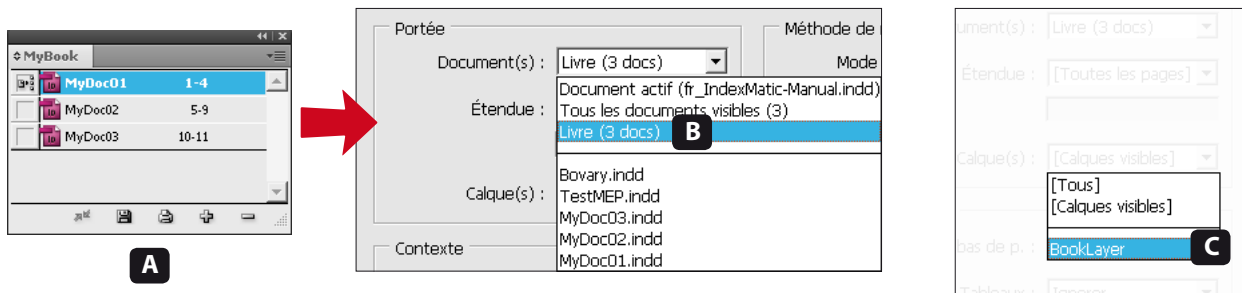
1. La liste déroulante **Document(s)** indique le(s) document(s) disponible(s) **B**. Le document actif est sélectionné par défaut. Vous pouvez en choisir un autre, ou même cibler tous les documents ouverts.
2. Si un seul document est ciblé, le champ **Étendue** **C** permet de restreindre les pages à explorer. Vous pouvez spécifier directement ces **Pages** (par ex. : **10-20**; **25**; **30-40**) ou choisir telle **section** du document. Par ailleurs, si un bloc-texte est actuellement sélectionné dans InDesign, vous pourrez cibler l'**article** associé.
3. Enfin, la liste **Calque(s)** **D** permet d'adresser un calque spécifique, ou seulement les calques **visibles**.

■ Lorsque vous ciblez **Tous les documents visibles**, gardez à l'esprit qu'aucune distinction ne sera faite entre des **folios identiques** pouvant apparaître dans des documents différents. Il vous appartient de configurer des **options de numérotation** pertinentes pour les documents que vous indexez. (Concernant l'indexation d'un livre → 5)

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Portée et contexte



5 INDEXATION D'UN LIVRE

Indexer un **livre entier** avec IndexMatic² est presque aussi simple qu'indexer un simple document :

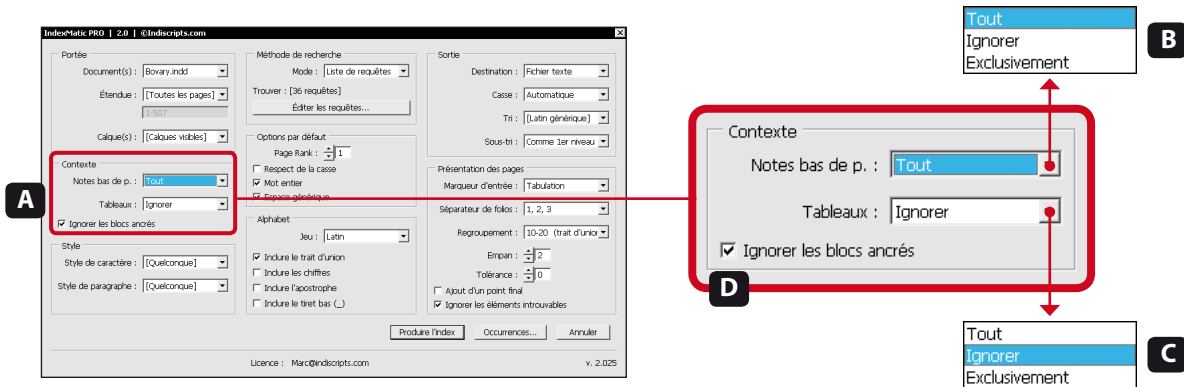
1. Ouvrez le livre dans InDesign **A**. (Il n'est pas nécessaire d'ouvrir les chapitres sous-jacents.)
2. Exécutez IndexMatic² et sélectionnez l'item « **Livre** » dans la liste des **Document(s)** **B**. Le nombre de documents associés est indiqué dans la liste. (Notez aussi que chaque document du livre apparaît séparément dans la liste, ce qui vous offre la possibilité d'indexer un chapitre particulier.)
3. Si besoin, ajustez le ou les **Calque(s)** cible(s) **C**. Lorsqu'un livre est sélectionné, la liste des calques n'affiche que ceux dont le nom appartient à **tous les documents** du livre — c.-à-d. les « calques communs ».

- IndexMatic² ne peut pas traiter plusieurs livres en même temps. Lors de l'indexation d'un livre, assurez-vous qu'**un seul fichier livre** est ouvert (i. e. un seul onglet) dans la palette d'InDesign.
- Si vous devez indexer **seulement quelques chapitres** plutôt que le livre entier, il est nécessaire d'ouvrir manuellement chaque fichier cible dans InDesign, puis de choisir la portée **Tous les documents visibles**.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Portée et contexte



6 AFFINER LE « CONTEXTE »

La rubrique **Contexte** **A** propose des filtres additionnels, agissant relativement à la portée courante :

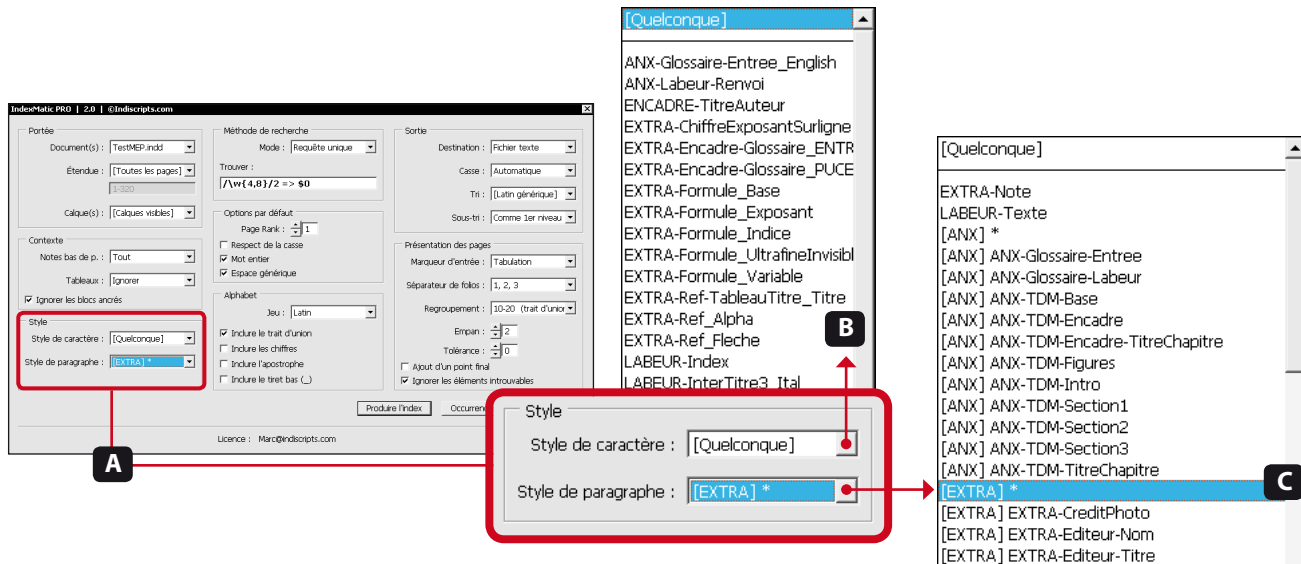
1. Le champ **Notes bas de p.** permet d'étendre / restreindre la recherche aux notes de bas de page. Dans la liste déroulante **B**, choisissez « Tout » pour inclure le contenu des notes, « Ignorer » pour l'exclure, ou « Seulement » pour restreindre spécifiquement le champ de recherche aux notes.
2. La même logique régit le champ **Tableaux** **C**, qui étend / restreint la recherche aux tableaux. Notez qu'IndexMatic² **n'explore que les cellules de niveau supérieur** (les sous-tableaux sont ignorés).
3. Cochez **Ignorer les blocs ancrés** **D** pour que le texte des blocs ancrés ou emboîtés ne soit pas pris en compte.

- Les notes de bas de page pouvant contenir des tableaux, il est possible — quoique restrictif ! — de combiner les options **Notes : Seulement** et **Tableaux : Seulement**.
- Les options de **Contexte** influent sur le temps d'exécution. Plus les filtres sont granulaires, plus le document est long à analyser.
- Avant la version 2.025, la localisation des **notes** et/ou **tableaux** était sujette à certaines approximations. Ces limitations sont révolues.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v.2.025

Portée et contexte



7 FILTRER SELON DES STYLES DE PARAGRAPHE / CARACTÈRE

Par nature, les styles de paragraphe et de caractère offrent au graphiste un contrôle total sur la présentation d'un document. Les styles permettent également de « baliser » des éléments sémantiques : titres, accroches, légendes, texte principal, descriptif de produit, etc. Grâce à la rubrique **Style A** d'IndexMatic², vous pourrez cibler sélectivement les parties de votre / vos document(s) via les styles sous-jacents :

1. **Style de caractère B** permet de cibler les contenus d'un style (ou groupe de styles) de **caractère**.
2. **Style de paragraphe C** permet de cibler les contenus d'un style (ou groupe de styles) de **paragraphe**.

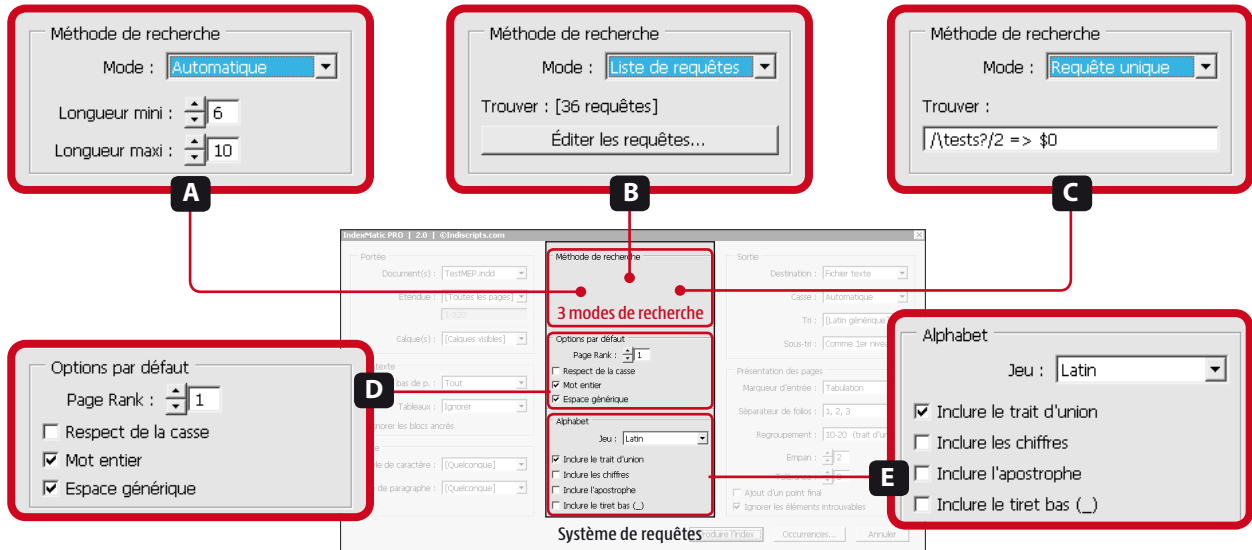
Les groupes de styles sont listés sous la forme **[groupe]*** (l'étoile signifiant : « tout style de ce groupe. »)

- Le filtrage par style(s) est disponible également lors de l'indexation d'un **livre** ou de **documents multiples**. Auquel cas, les listes de styles n'indiquent que les éléments **communs** aux documents-cibles.
- Le moteur de recherche d'IndexMatic² considère les styles appliqués **et non pas les enrichissements** du texte. Il s'ensuit que les reformatages locaux sont sans effet sur le processus de filtrage.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de recherche



8

INTRODUCTION AU SYSTÈME DE REQUÊTES

Une fois que vous avez défini la partie du/document(s) soumis à l'indexation (*v. pages précédentes*), vous devez indiquer à IndexMatic² comment extraire les mots pertinents, sous quelles formes et conditions.

1. La rubrique **Méthode de recherche** possède trois états : **Automatique**, **Liste de requêtes** et **Requête unique**. **Automatique** **A** permet de collecter tous les mots de longueur délimitée. **Liste de requêtes** **B** envoie au moteur un ensemble d'instructions (liste de mots, expressions régulières...). Enfin, **Requête unique** **C** permet de tester rapidement une simple requête.
2. Les rubriques **Options par défaut** **D** et **Alphabet** **E** regroupent les paramètres généraux du système de requêtes. Ils s'appliquent à tous les modes de recherche et précisent comment « interpréter » les requêtes.

■ Lors de votre session InDesign, IndexMatic² PRO mémorise les derniers réglages effectués dans la boîte de dialogue (portée, contexte, styles, requêtes, etc.). Ainsi, vous pouvez affiner progressivement vos paramètres à chaque nouvelle exécution du script.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de recherche

Eribusandiat omni autatendit quunt fuga. Peirce rit porersped quod es-tecuptis doluptata-tem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem vollesitate prem-pem utem con Uci-	picient vendis do-lorehenis as Peirce aut offic tem conse-quam rehendit aut quiaie aut expe pel-liam rent carum sant Peirce excec idu-citissam idi dolorio qui saepa eiu? Ci-ligent. Eprepre ve-
20	21

Eribusandiat omni autatendit quunt fuga. Peirce rit porersped quod es-tecuptis doluptata-tem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem vollesitate prem-pem utem con Uci-	picient vendis do-lorehenis as Peirce aut offic tem conse-quam rehendit aut quiaie aut expe pel-liam rent carum sant Peirce excec idu-citissam idi dolorio qui saepa eiu? Ci-ligent. Eprepre ve-
20	21

Eribusandiat omni autatendit quunt fuga. Peirce rit porersped quod es-tecuptis doluptata-tem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem vollesitate prem-pem utem con Uci-	picient vendis do-lorehenis as Peirce aut offic tem conse-quam rehendit aut quiaie aut expe pel-liam rent carum sant Peirce excec idu-citissam idi dolorio qui saepa eiu? Ci-ligent. Eprepre ve-
20	21

Eribusandiat omni autatendit quunt fuga. Peirce rit porersped quod es-tecuptis doluptata-tem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem vollesitate prem-pem utem con Uci-	picient vendis do-lorehenis as Peirce aut offic tem conse-quam rehendit aut quiaie aut expe pel-liam rent carum sant Peirce excec idu-citissam idi dolorio qui saepa eiu? Ci-ligent. Eprepre ve-
20	21

nimpo iciunt et etur mo blaut et is ex et harupti quam illa-estiae niae alit mod mos et ex eum none vellatincis molesto voluptaspis cum et et, ut mo voliecti apeditam quia nese-quaeh rehenti squisit	Peirce sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni-dus nos unti Peirce corem rendae non-sequid enis utem in estis saepa eiu, co-rero Peirce
22	23

nimpo iciunt et etur mo blaut et is ex et harupti quam illa-estiae niae alit mod mos et ex eum none vellatincis molesto voluptaspis cum et et, ut mo voliecti apeditam quia nese-quaeh rehenti squisit	Peirce sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni-dus nos unti Peirce corem rendae non-sequid enis utem in estis saepa eiu, co-rero Peirce
22	23

nimpo iciunt et etur mo blaut et is ex et harupti quam illa-estiae niae alit mod mos et ex eum none vellatincis molesto voluptaspis cum et et, ut mo voliecti apeditam quia nese-quaeh rehenti squisit	Peirce sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni-dus nos unti Peirce corem rendae non-sequid enis utem in estis saepa eiu, co-rero Peirce
22	23

nimpo iciunt et etur mo blaut et is ex et harupti quam illa-estiae niae alit mod mos et ex eum none vellatincis molesto voluptaspis cum et et, ut mo voliecti apeditam quia nese-quaeh rehenti squisit	Peirce sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni-dus nos unti Peirce corem rendae non-sequid enis utem in estis saepa eiu, co-rero Peirce
22	23

Page Rank=1
Peirce 20, 21, 23

Page Rank=2
Peirce 21, 23

Page Rank=3
Peirce 23

Page Rank>3
Peirce —
(non trouvé)

10 SIGNIFICATION DU « PAGE RANK »

Dans IndexMatic², le **page rank** est une valeur de 1 à 9 qui représente le **nombre minimum d'occurrences** d'une expression donnée dans une page donnée. Le principe sous-jacent est le suivant : « plus un mot apparaît fréquemment dans une page, plus il est pertinent d'associer cette page à ce mot dans l'index final ».

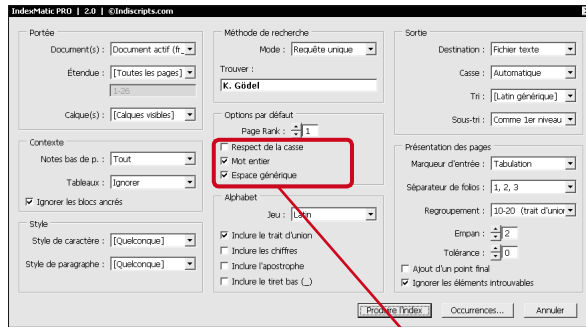
1. Lorsque vous fixez un page rank supérieur à 1 — disons 3 — dans la rubrique **Options par défaut**, vous imposez au module de recherche de négliger toute expression apparaissant moins de 3 fois dans une page.
2. Augmenter le page rank est tout particulièrement recommandé en mode de recherche **Automatique** (lequel tend souvent à produire un trop grand nombre de mots).

- Tous les **modes de recherche** sont soumis au page rank. Ce dernier agit globalement, à moins qu'une requête particulière inhibe ou supplante la valeur par défaut.
- Même si une expression a été « trouvée » dans le texte, mais sans jamais passer le test du page rank, elle sera réputée **non trouvée** et éventuellement reportée comme telle → 26.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de recherche



A

- Respect de la casse
- Mot entier
- Espace générique

B

Concordance de l'espace générique

Cette table liste tous les caractères reconnus par l'espace simple lorsque l'option « Espace générique » est activée.

* Le Saut de ligne forcé a été introduit dans la v. 2.025

Nom dans InDesign	GREP	Unicode	Symb.
Espace		U+0020	·
Saut de ligne forcé*	\n	U+000A	↵
Espace insécable justifiante	~S	U+00A0	^
Espace insécable à chasse fixe	~S	U+202F	^
Tabulation	\t	U+0009	»
Retrait jusqu'à ce point	~i	U+0007	↑
Tabulation de retrait à droite	~y	U+0008	↵
Espace sans alinéa	~f	U+2001	~
Espace demi-cadratin	~>	U+2002	—
Espace cadratin	~m	U+2003	—
Tiers d'espace	~3	U+2004	·
Quart d'espace	~4	U+2005	·
Sixième d'espace	~%	U+2006	·
Espace tabulaire (figure space)	~/	U+2007	#
Espace de ponctuation	~.	U+2008	!
Espace fine	~<	U+2009	·
Espace ultra-fine	~	U+200A	·
Espace moyenne mathématique		U+205F	↵

11

OPTIONS DE CONCORDANCE

Les **Options par défaut** contiennent trois cases à cocher **A** qui permettent de renforcer ou d'atténuer certaines contraintes dites « de concordance ». Mettons que nous recherchions les occurrences de la chaîne **K. Gödel** :

1. En activant **Respect de la casse**, nous indiquons au système de ne considérer que les concordances exactes et de rejeter toute variante telle que **K. GÖDEL**. (Cette contrainte est toutefois désactivée par défaut.)
2. L'option **Mot entier** (cochée par défaut) impose qu'une concordance ne soit pas inscrite dans un mot plus long. Plus exactement, aucun **caractère de l'alphabet actif** (→9) ne doit précéder ou suivre l'expression.
3. L'option **Espace générique** (cochée par défaut) indique que l'espace simple située entre **K.** et **Gödel** peut concorder avec **tout autre caractère « blanc » disponible dans InDesign** (insécable, fine, tabulation... **B**).

- Le **Respect de la casse** intéresse les caractères eux-mêmes (majuscules vs minuscules) et ne tient pas compte des effets de formatage. Par exemple, même si de petites capitales sont appliquées à une minuscule, le moteur de recherche continue de « voir » ce caractère comme une minuscule.
- L'option **Mot entier** n'interdit pas que l'expression recherchée comporte elle-même des espaces ou autres caractères absents de l'alphabet.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de recherche

Méthode de recherche
Mode : Automatique **A**

Longueur mini : 6 **B**
Longueur maxi : 10

Options par défaut
Page Rank : 3 **C**
 Respect de la casse
 Metastyles

Style
Style de caractère : [Quelconque]
Style de paragraphe : Main-Text **D**



10	billard	95
11	bouquet	41
12	Bovary	13, 54, 117, 247, 268
13	Bridoux	401
14	chaque	229
15	Charbovari	4
16	Charles	9, 13, 15, 18, 24, 30-31
17	chevaux	221
18	classe	4
19	devant	68, 72, 119, 346
20	devoir	197
21	disait	479
22	docteur	467
23	dormait	494
24	église	337, 489
25	encore	369
26	enfant	115
27	étaient	19, 70
28	francs	297, 388
29	habits	33
30	heures	96
31	Hippolyte	249
32	Homais	92, 112, 117, 127, 155,

12 UTILISATION DU MODE « AUTOMATIQUE »

Le mode de recherche **Automatique** permet typiquement d'extraire le **vocabulaire** d'un ouvrage, ou encore d'indexer les **noms de produits** d'un catalogue (sous réserve que ces noms soient correctement stylés).

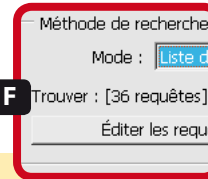
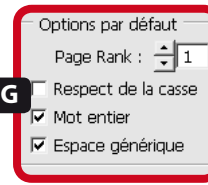
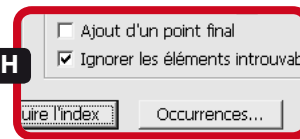
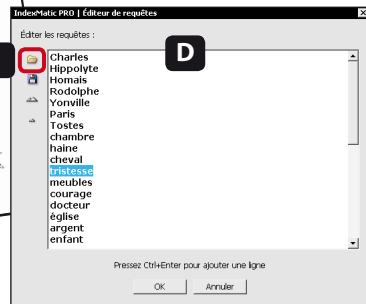
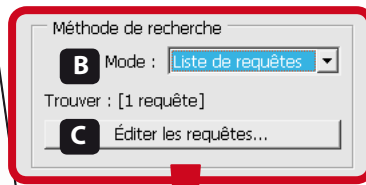
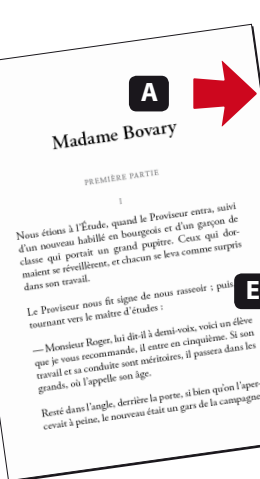
1. Dans la liste déroulante du champ **Mode**, sélectionnez « Automatique » **A**.
2. Ajustez la **Longueur mini** (≥ 2) et la **Longueur maxi** (≤ 40) **B**. Le moteur de recherche extraira alors toute séquence de caractères ainsi calibrée, et formant un « mot entier » au sens de l'**alphabet actif** →9.
3. Dans l'hypothèse où vous collectez le vocabulaire d'un ouvrage, songez à augmenter le **page rank** **C** afin d'extraire les termes les mieux représentés. Le cas échéant, appliquez un filtre de **style** **D** →7.

- En mode Automatique, les **options de concordance** sont inactives. En revanche, le **page rank** est actif.
- À noter que les expressions contenant des **espaces internes** sont par nature inaccessibles à la recherche automatique. Si vous avez besoin d'indexer des chaînes quelconques formées dans un style de caractère particulier, le mieux est de lancer une **requête regex** →19, telle que : `/+/`

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de recherche



1 argent	6, 13, 19, 24, 48, 61, 82, 88
2 chambre	9-11, 25, 28, 41, 43, 58, 72,
3 Charles	4, 9, 13-26, 28-32, 35, 38-42,
4 cheval	11, 14, 16, 21, 31-32, 42, 47,
5 couloir	348
6 courage	211, 220, 251, 270, 283, 415,
7 docteur	215, 231, 243, 253-255, 258, 2
8 église	8, 23, 34, 39, 49, 59, 82-83,
9 enfant	7-10, 16, 43, 62, 115, 117, 11
10 étage	36, 92, 106, 193, 207, 287-288
11 fenêtre	11, 19-20, 24, 27, 29, 31, 41-
12 figure	5, 12, 22, 27, 29, 53, 60, 64-
13 haïne	141, 143, 167, 437, 499
14 Hippolyte	191, 242-250, 252, 254-256, 26
15 Homais	92, 94-95, 97, 99, 102-103, 16
16 lettre	10, 14-15, 49, 158, 226, 238-2
17 maison	6-7, 12, 16-17, 24, 29, 38-39,
18 malheur	24-25, 27, 101, 217, 238, 278,
19 marchand	6, 45, 93, 101, 112, 134, 263,
20 meubles	10, 27, 65, 68, 78, 110, 153,
21 nourrice	7, 101, 118-123, 146, 230, 302
22 Paris	51, 62, 73-75, 77, 91, 101, 10
23 pauvre	3-4, 10, 23, 25, 27, 45, 71, 7
24 pharmacien	87, 91-100, 104, 106-107, 109,
25 regard	18, 28, 51, 122-123, 136, 151,
26 Rodolphe	171, 176-177, 184-190, 193-197
27 science	104, 102, 244, 248, 392, 396,
28 soleil	11, 17, 22, 27, 42, 48, 50, 58
29 sommeil	15, 194, 199
30 toit	59, 91, 142, 214, 226, 274, 29
31 Tostes	13, 15, 24, 39, 58, 70, 73, 76
32 tristesse	40, 61, 84, 122, 142, 167, 188

13

CRÉER UN INDEX À PARTIR D'UNE LISTE DE MOTS

Vous venez de finir la mise en page de *Madame Bovary* **A**. En vue d'une recherche universitaire, l'éditeur souhaite à présent établir un index à partir d'une liste de termes prédéfinis : *haine, courage, tristesse, Yonville...*

1. Exécutez IndexMatic² et choisissez « Liste de requêtes » **B** dans la rubrique **Méthode de recherche**.
2. L'**Éditeur de requêtes** **D** s'ouvre automatiquement (sinon, cliquez sur **Éditer les requêtes...** **C**). Ici vous pouvez saisir une liste de mots-clés, ou bien l'importer depuis un fichier **E** grâce au bouton → **14**
3. Pressez **OK** pour fermer l'éditeur. Le panneau de recherche indique alors le **nombre de requêtes** enregistrés **F**. Ajustez les **options de concordance**, le **page rank** **G** et autres options de sortie **H**. Quand tout est paramétré selon vos besoins, cliquez sur le bouton **Produire l'index**.

- Lorsque vous soumettez à IndexMatic² une liste de mots traditionnelle, il est préférable d'activer l'option **Mot entier** afin d'empêcher des **concordances partielles** telles que : *bar*be, *bar*que, *bar*atin...
- Les items d'une liste de mots ne subissent pas les limites de l'**Alphabet** actif → **9**. Y sont autorisés les caractères exotiques, les espaces, etc. En cela, une liste de « mots » permet de produire des index de toute nature.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de recherche

Ouvrir / sauvegarder une liste. **B**

Augmenter / réduire la taille des caractères. **C**

A

Ces commentaires ne seront pas analysés comme des requêtes.

Requêtes simples basées sur des VOCABLES. →16

Requêtes basées sur des MOTIFS. →19

Un motif un peu plus complexe. →19

Pressez Ctrl+Enter pour ajouter une ligne

OK **D** Annuler

14 UTILISER L'ÉDITEUR DE REQUÊTES

L'**Éditeur de requêtes** est un module accessible en mode « Liste de requêtes » via le bouton « Éditer les requêtes... ». Il permet d'éditer et de gérer une liste de requêtes →15 destinée au moteur de recherche.

1. Vous pouvez librement saisir, couper/copier/coller des éléments dans la zone d'édition principale **A**. (Certaines plateformes réclament la combinaison **Ctrl+Entrée** pour produire les sauts de ligne.)
2. Cliquez sur pour ouvrir un fichier texte. Cliquez sur pour sauvegarder la liste dans un fichier texte. **B**
3. Pour augmenter ou réduire la taille du texte **C** dans la zone d'édition, cliquez respectivement sur ou .
4. Pressez **OK** pour fermer l'éditeur en prenant en compte les changements opérés. Pressez **Annuler** pour ignorer les changements (la liste est alors restaurée dans son état précédent et l'éditeur refermé). **D**

■ Depuis IndexMatic v. 2.025, une ligne débutant par deux barres obliques (//) est traitée comme un simple **commentaire**. Les commentaires ne sont pas « interprétés » et permettent de rendre la liste plus lisible.

■ Les commentaires de la forme :
// sujet1 => sujet2
indiquent des références croisées →20

■ Toute **ligne vide** est ignorée et supprimée de la liste.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées

R E Q U Ê T E

ce que vous
CHERCHEZ



ce que vous
INDEXEZ

A

B

bateau

=>

bateau

Terme implicite
(non requis)

Gödel
navire
barque

=>

Gödel, K.
bateau
bateau

Terme explicite
(réécriture ou
regroupement)

15

SYNTAXE DES REQUÊTES (PRÉAMBULE)

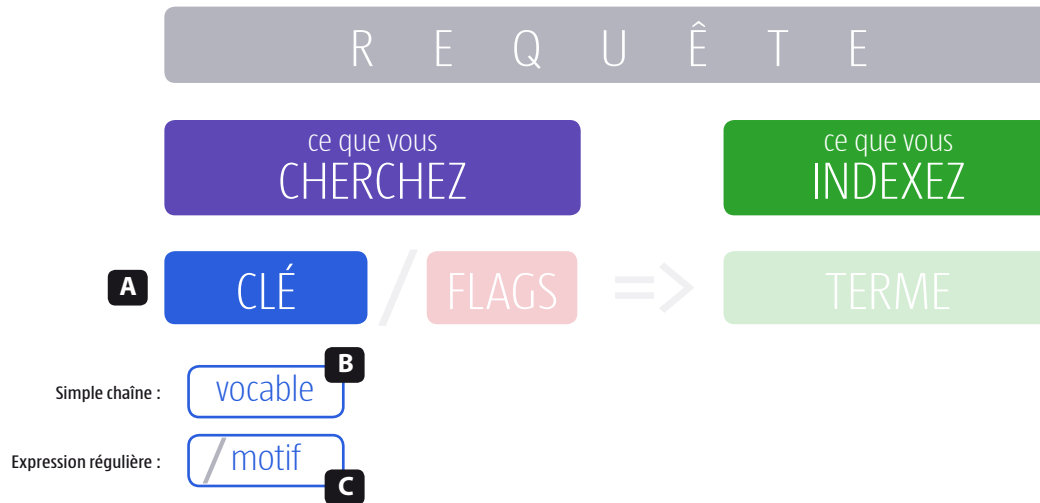
1. Ce qui rend IndexMatic² extraordinairement flexible est son **interpréteur** de requêtes. Chaque requête est scindée en deux parties : ce qui est **RECHERCHÉ** **A**, ce qui est **INDEXÉ** **B**. Le plus souvent, cette dichotomie reste implicite et invisible à l'utilisateur. Par exemple, lorsque vous soumettez un simple mot, disons **bateau**, le système présume que **bateau** est à la fois l'expression cherchée et le terme à reporter dans l'index. En tâche de fond, cette requête est interprétée ainsi : **bateau=>bateau**.
2. Mais ce mécanisme peut se révéler insuffisant. D'une part, on pourrait souhaiter **réécrire** un terme : **Gödel=>Gödel, K.** D'autre part, on pourrait souhaiter capturer les **variantes** ou **regrouper** plusieurs mots (*navire, barque*) sous un **seul sujet** (*bateau*). Pour ce faire vous utiliserez l'opérateur => comme illustré ci-après.

- L'interpréteur syntaxique travaille aussi bien en mode **Liste de requêtes** qu'en mode **Requête unique**.
- L'**opérateur de réécriture** doit toujours être saisi ainsi : => (signe d'égalité suivi du signe « supérieur à »). Vous pouvez sans risque ajouter des **espaces intercalaires** entre tous les opérateurs syntaxiques.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées



16

SYNTAXE DES REQUÊTES : LA « CLÉ »

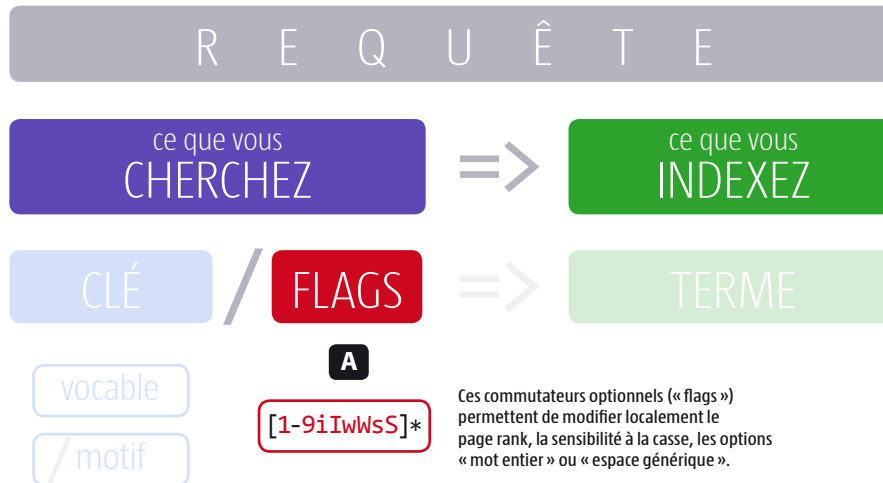
1. La **CLÉ** **A** est le **seul élément obligatoire** d'une requête. Il s'agit le plus souvent du mot ou de l'expression à rechercher — que nous appelons alors un **vocable** **B**. Tant que vous n'utilisez pas d'opérateurs syntaxiques, toute chaîne de caractères est analysée comme un vocable et forme en soi une requête valide et complète.
2. On peut aussi fournir une **expression régulière**. Pour ce faire, il suffit d'amorcer le **motif** **C** de la regex par une barre oblique (/). Exemples : **/chien|chat** trouvera l'un de ces deux mots, **/chats?** trouvera *chat* et *chats*. Il est possible mais non nécessaire de clore la regex par une autre barre oblique : **/chats?/** →**19**.
3. En fonction des **Options par défaut** →**11**, une CLÉ (vocable ou motif) peut être **sensible ou insensible** à la casse. De même, elle supportera ou non **l'espace générique** ou autres options similaires →**17**.

- La barre oblique / ayant une signification particulière au sein d'une CLÉ, vous devrez utiliser la séquence d'échappement \ si vous avez besoin d'introduire ce caractère en tant que tel.
- Réciproquement, tout motif doit s'amorcer par une barre oblique. À défaut, il serait vu comme un pur **vocable**, ce qui rendrait bien sûr inopérant tout métacaractère propre aux expressions régulières.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées



17 SYNTAXE DES REQUÊTES : LES « FLAGS »

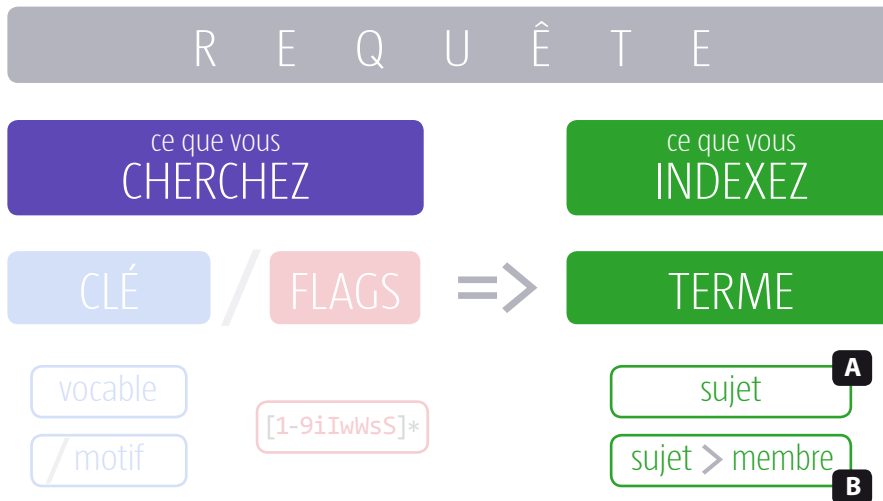
1. Nous appelons **FLAGS** **A** un jeu de commutateurs optionnels agissant localement sur les options de concordance. Une séquence de flags doit s'amorcer par une barre oblique (/). Les flags disponibles sont : **1 2 3 4 5 6 7 8 9 i I w W s S**. On peut les combiner dans un ordre arbitraire : **chat/3w,/USA|ÉU/I2**.
2. Un chiffre de **1** à **9** est interprété comme flag de **page rank**, il écrase donc la valeur par défaut → **10**.
3. De même, quelles que soient les options par défaut :
 - i** rend la requête **insensible à la casse**, **I** la rend **sensible à la casse** → **11**;
 - w** active l'option **mot entier**, **W** inhibe l'option **mot entier** → **11**;
 - s** active l'**espace générique**, **S** désactive l'**espace générique** → **11**.

- Les **FLAGS** servent à **supplanter** ou **inhiber** les options par défaut lorsque c'est nécessaire. Par exemple, si vous avez configuré le page rank à 3, vous pouvez atténuer localement ce critère afin de formuler une requête moins drastique : **abracadabra/1**
- Notez que les flags **s** et **S** sont sans effet si la CLÉ ne contient pas d'espace.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées



Expliciter le TERME ouvre de nombreuses perspectives : grouper plusieurs clés sous un même sujet, traduire un motif de regex en un terme lisible, déclarer la relation hiérarchique entre un sujet et ses membres.

18 SYNTAXE DES REQUÊTES : LE « TERME »

Le **TERME** (à droite du modèle) reflète la chaîne de caractères à insérer finalement dans l'index. En fonction de la CLÉ, l'interpréteur peut déterminer implicitement un ou plusieurs TERME(s) par défaut. Cependant, il reste possible de **découpler formellement** le TERME et la CLÉ via l'opérateur de **réécriture** (=>).

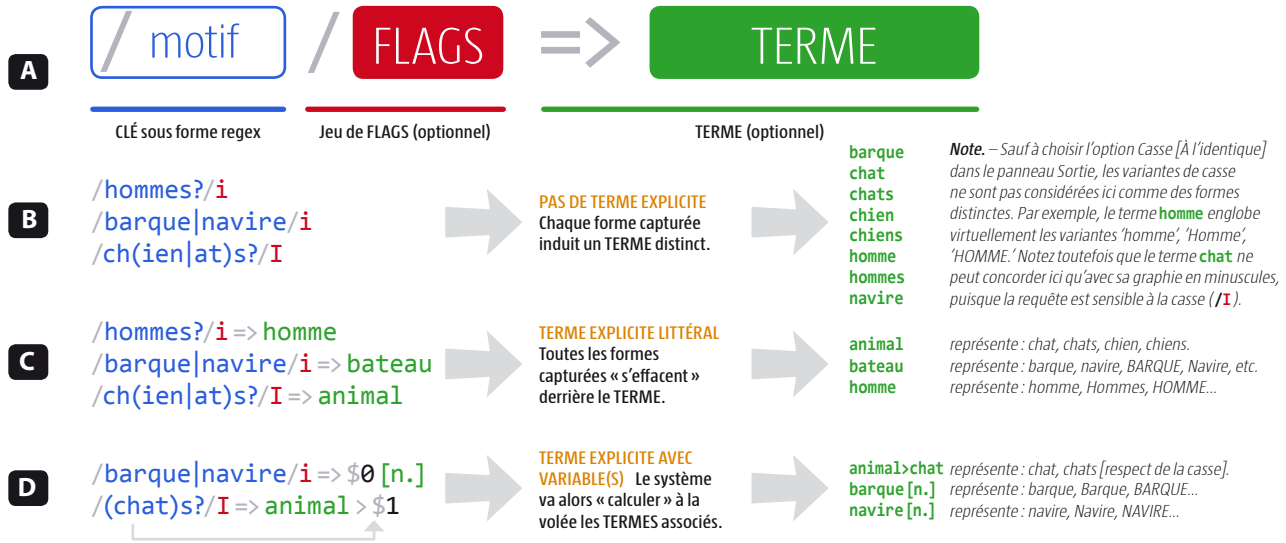
1. Dans sa forme atomique, un TERME est une simple chaîne littérale, par ex. **animal**. Il prend la place de la CLÉ et tient lieu de sujet (*topic*) dans l'index **A**. Ainsi: **chat=>animal** demande à l'interpréteur de connecter le vocable **chat** au sujet **animal**. (On peut bien sûr réemployer un même TERME dans plusieurs requêtes.)
2. La syntaxe d'un TERME admet aussi un opérateur formel (>) qui déclare un élément comme *membre* d'un sujet **B**. Par exemple, **chat=>animal>chat** indexe **chat** comme membre du sujet **animal**.

- Représentez-vous un sujet comme un terme de premier niveau qui, le cas échéant, peut **chapeauter** des termes de second niveau (membres). IndexMatic² n'admet que **ces deux échelons**.
- Le symbole > jouant un rôle particulier dans la syntaxe d'un terme, vous devez utiliser la séquence >> pour introduire ce caractère en tant que tel.
- Concernant la **mise en forme finale** des termes d'index : ->23 et suiv.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées



19

AJUSTER LES REQUÊTES BASÉES SUR UNE REGEX

- Le propre d'une expression régulière (regex) est de capturer **différentes formes** dans le texte. Lors du traitement d'une requête **basée sur une regex** **A**, IndexMatic² présume que **chaque forme** concordante doit produire un TERME distinct **B**, à moins qu'un TERME explicite (littéral) soit indiqué.
- Si la requête possède un TERME **explicite**, alors l'opérateur de réécriture agrège sous ce TERME — sujet ou membre — les différentes formes capturées. Ainsi, le TERME « efface » la cuisine interne de la regex **C**.
- Enfin, la syntaxe d'un TERME autorise l'insertion de **variables** (**\$0**, **\$1**, ..., **\$9**). Celles-ci représentent les **sous-chaînes** issues de **parenthèses capturantes**. **\$0** représente la capture globale, tandis que **\$n** représente la sous-chaîne de rang *n* au sein de la regex (selon le décompte des parenthèses ouvrantes) **D**.

- Grâce à l'option « Casse » de la rubrique Sortie, les termes de l'index peuvent être « reformattés » *a posteriori* => **24**
- Les versions préliminaires du script autorisaient l'usage du symbole **\$** comme abréviation de **\$0**. Désormais, ce symbole employé seul représente la **CLÉ elle-même** (vocabulaire ou motif). Ainsi, **/chats?/ => \$** équivaut à : **/chats?/ => chats?**

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées



B // bateau ⇒ Voir aussi navire.
// animal > chat ⇒ Voir: félidés.

20

INDICATION DE RÉFÉRENCES OU RENVOIS

1. Il est possible de simuler un renvoi **A** grâce à la syntaxe : // TERME ⇒ RÉFÉRENCE
Une « requête » ainsi rédigée n'est en fait qu'un simple commentaire, mais elle impose au moteur de créer un PSEUDO-TERME dans l'index et de lui accoler le texte de RÉFÉRENCE, ce qui permet de formuler un renvoi vers un autre terme, ou toute autre indication. (Aucune clé « réelle » n'est autorisée ici.)
2. Tout TERME syntaxiquement valide est admissible **B**, y compris sous la forme **sujet> membre →18**, mais naturellement vous ne pouvez pas utiliser de variable **\$n** dans ce contexte.
3. Toute chaîne de caractères peut servir de RÉFÉRENCE : « Voir ... », « Voir aussi : ... », etc. L'interpréteur ne présume ni ne contrôle l'existence éventuelle du (ou des) terme(s) qui seraient mentionnés ici.

- Une consigne de renvoi n'est pas effectivement adressée au module de recherche. Le pseudo-terme est seulement traité « comme si » il avait résulté d'une requête normale, et sa référence est alors affichée en lieu et place des folios →23.
- Le marqueur de commentaire (//) et l'opérateur ⇒ sont l'un et l'autre nécessaires pour créer un renvoi. Si l'opérateur ⇒ était omis, la « requête » serait analysée comme un simple commentaire →14.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées

Opérateur	Usage
<code>\</code>	Caractère d'échappement devant opérateurs : <code>. \ / + * ? [^ \$ () { } = ! < > : - </code>
<code>*</code>	Capture l'élément antéposé 0 ou plusieurs fois.
<code>+</code>	Capture l'élément antéposé 1 ou plusieurs fois.
<code>?</code>	Capture l'élément antéposé 0 ou 1 fois.
<code>{n}</code>	Capture exactement <i>n</i> occurrences de l'élément antéposé.
<code>{n,}</code>	Capture au moins <i>n</i> occurrences de l'élément antéposé.
<code>{n,m}</code>	Capture au moins <i>n</i> occurrences et au plus <i>m</i> occurrences de l'élément antéposé.
<code>x y</code>	Alternative. Capture soit <i>x</i> , soit <i>y</i> .
<code>(x)</code>	Parenthèses capturantes. (Voir \$ → 19.)
<code>(?:x)</code>	Parenthèses non capturantes.
<code>x(?:y)</code>	« Lookahead » positif.
<code>x(?:!y)</code>	« Lookahead » négatif.
<code>[xyz]</code>	Classe de caractères. Capture n'importe quel caractère spécifié entre les crochets.
<code>[^xyz]</code>	Classe complémentaire. (Le symbole <code>^</code> n'est pas utilisé dans d'autres contextes.)

(n) Les caractères suivants étant supprimés du flux de recherche, ils ne sont *capturés par aucun* métacaractère : saut de paragraphe, saut de colonne, saut de bloc, saut de page, numéro de page, variable de texte, Fin de style imbriqué, glyphe spécial « Non Roman », trait d'union conditionnel, saut de ligne conditionnel, « Non-joiner », « Zero Width Joiner », et tout marqueur ou ancre.

Attention : `\w`, `\W`, `\l`, `\L`, `\m` et `\M` sont tous *relatifs à l'alphabet courant*, ce qui n'est pas leur sémantique conventionnelle. Notez d'autre part que le métacaractère adressant les majuscules est `\m` et `\M` (et non pas `\u` et `\U`).

Les classes et classes complémentaires autorisent les métacaractères : `\w`, `\W`, `\l`, `\L`, `\m`, `\M`, `\uHHHH`, ainsi que l'ESPACE (si espace générique).

Métacar.	Usage
ESPACE	Si l'option Espace générique est activée, capture tout caractère « blanc » (v. tableau → 11.) Séquence d'échappement : <code>\ESPACE</code>
<code>.</code> (point)	Capture n'importe quel caractère ⁽ⁿ⁾ .
<code>\d</code>	Capture tout chiffre. Équivalent à <code>[0-9]</code> .
<code>\D</code>	Capture tout non-chiffre, i. e. <code>[^0-9]</code> .
<code>\w</code>	Capture tout caractère présent dans l'alphabet courant (Voir → 9.)
<code>\W</code>	Capture tout caractère absent de l'alphabet courant. (Voir → 9.)
<code>\l</code>	Capture toute lettre minuscule de l'alphabet courant. (Voir → 9.)
<code>\L</code>	Capture tout caractère qui n'est pas une lettre minuscule de l'alphabet courant. (Voir → 9.)
<code>\m</code>	Capture toute lettre majuscule de l'alphabet courant. (Voir → 9.)
<code>\M</code>	Capture tout caractère qui n'est pas une lettre majuscule de l'alphabet courant. (Voir → 9.)
<code>\t</code>	Capture le caractère de tabulation.
<code>\s</code>	Équivalent à : <code>[\t\u00A0\u2028\u2029]</code> (moins efficace en général que l'espace générique)
<code>\uHHHH</code>	Capture le caractère Unicode U+HHHH.
<code>\b</code> <code>\B</code> <code>\s</code> <code>\S</code> <code>\f</code> <code>\n</code> <code>\r</code>	<i>Inusités dans les motifs.</i>

■ Les symboles `^` et `$` sont sans usage dans les motifs. Par contre, au sein d'un TERME, `$` représente la clé originale (`$`) ou une variable capturée (`$0`, `$1`, `$2`, ..., `$$`). Séquence d'échappement : `\$`.

21

OPÉRATEURS SPÉCIAUX DANS LES REGEX

- IndexMatic² supporte les **expressions régulières ECMA Script (RegExp)**, telles que décrites dans la spécification **ECMA-262** (section 15.10) : <http://www.ecma-international.org/publications/standards/Ecma-262.htm> Un excellent aperçu de cette syntaxe, assorti d'exemples pratiques, est disponible (en anglais) à cette adresse : http://developer.mozilla.org/en/JavaScript/Guide/Regular_Expressions (MDN's Guide to Regular Expressions).
- En appoint, IndexMatic² ajoute ou redéfinit certains opérateurs **A** de façon à moduler ou amender la spécification officielle. Ces derniers simplifient la syntaxe des requêtes ou permettent des captures plus fines. (Voir aussi page suivante.)

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Requêtes avancées

Attention : lorsque vous utilisez un métacaractère qui discrimine la casse, comme `\p{Lu}`, assurez-vous que votre requête est elle-même sensible à la casse → **11**

Bien que le métacaractère « négatif » `\p{...}` ne soit pas implémenté, il reste possible d'utiliser une classe complémentaire de la forme : `[^\p{xx}]`.

A Propriétés Unicode

Propriété	Usage	Propriété	Usage
<code>\p{Ll}</code>	Lettre minuscule. Ex. : é, α, ā, œ, ç...	<code>\p{Ps}</code>	Ponctuation ouvrante. Ex. : (, [, {...
<code>\p{Lu}</code>	Lettre majuscule. Ex. : A, É, Ð, Ω, NJ...	<code>\p{Pe}</code>	Ponctuation fermante. Ex. :),], }...
<code>\p{Lt}</code>	Lettre en casse de titre. Ex. : Dž, Lj, Nj...	<code>\p{Pi}</code>	Ponctuation initiale. Ex. : «, ’, ”...
<code>\p{Lm}</code>	Lettre modificatrice. Ex. : ħ, ı, ħ...	<code>\p{Pf}</code>	Ponctuation finale. Ex. : »; ’, ”...
<code>\p{Lo}</code>	Lettre sans casse. Ex. : 2, ll, ă...	<code>\p{Pc}</code>	Connecteur. Ex. : _; , { ...
<code>\p{M}</code>	Toute marque.	<code>\p{Po}</code>	Autre ponctuation. Ex. : !; ; #...
<code>\p{Mn}</code>	Marque à chasse nulle.	<code>\p{S}</code>	Tout symbole.
<code>\p{Mc}</code>	Marque combinatoire avec chasse.	<code>\p{Sm}</code>	Symbole mathématique. Ex. : +, <, ↔...
<code>\p{Me}</code>	Marque englobante.	<code>\p{Sc}</code>	Symbole monétaire. Ex. : \$, €, £, ₪...
<code>\p{N}</code>	Tout nombre.	<code>\p{Sk}</code>	Symbole modificateur. Ex. : ^, ¯, ˇ, ...
<code>\p{Nd}</code>	Chiffre décimal. Ex. : 1, 2, ε...	<code>\p{So}</code>	Autre symbole. Ex. : †, §, ©...
<code>\p{Nl}</code>	Lettre numérale. Ex. : viii, @, %...	<code>\p{Z}</code>	Tout séparateur.
<code>\p{No}</code>	Autre nombre. Ex. : ², ¼, ı...	<code>\p{Zs}</code>	Espace.
<code>\p{P}</code>	Tout signe de ponctuation.	<code>\p{C}</code>	Tout autre caractère.
<code>\p{Pd}</code>	Tiret et assimilé. Ex. : -, —, ~...		

B

Raccourcis GREP additionnels

Méta	Caractère
<code>~8</code>	Puce •
<code>~e</code>	Points de suspension ...
<code>~7</code>	Pied-de-mouche ¶
<code>~6</code>	Symbole de section §
<code>~2</code>	Copyright ©
<code>~r</code>	Marque déposée ®
<code>~d</code>	Marque commerciale ™
<code>~_</code>	Tiret cadratin —
<code>~=</code>	Tiret demi-cadratin –
<code>~{</code>	Guillemets ouvrants “
<code>~}</code>	Guillemets fermants ”
<code>~[</code>	Guillemet simple ouv. ‘
<code>~]</code>	Apostrophe ’
<code>~”</code>	Guillemets dactylo. ”
<code>~'</code>	Apostrophe dactylo. '
<code>~S</code>	Esp. insécable justifiante
<code>~s</code>	Esp. insécable chasse fixe
<code>~i</code>	Retrait jusqu'à ce point
<code>~y</code>	Tabulation à droite
<code>~f</code>	Espace sans alinéa
<code>~></code>	Espace demi-cadratin
<code>~m</code>	Espace cadratin
<code>~3</code>	Tiers d'espace
<code>~4</code>	Quart d'espace
<code>~%</code>	Sixième d'espace
<code>~/</code>	Espace tabulaire
<code>~.</code>	Espace de ponctuation
<code>~<</code>	Espace fine
<code>~ </code>	Espace ultra-fine
<code>~~</code>	Tilde (échappement)

■ Gardez à l'esprit que la sémantique des expressions régulières d'IndexMatic² ne coïncide pas en tous points avec celle de GREP InDesign (en dépit de très fortes similarités).

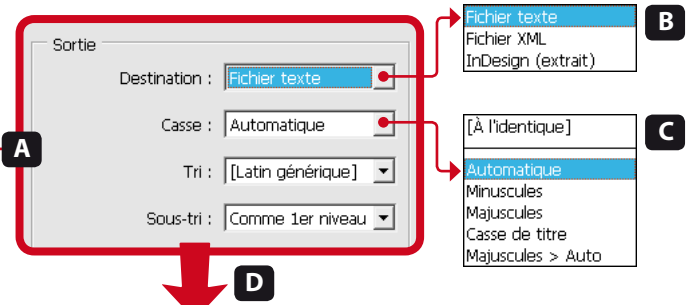
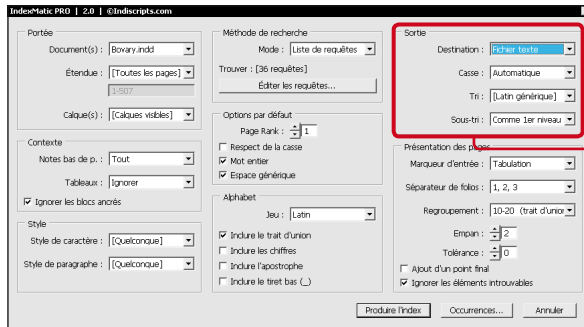
22 MÉTACARACTÈRES ADDITIONNELS

- Depuis la v. 2.025, IndexMatic² prend en charge le métacaractère `\p` sous la forme `\p{xx}`, où `xx` désigne une **propriété Unicode**. Les propriétés disponibles sont décrites dans la table **A**. Cette fonction permet de recherche des caractères par catégorie **indépendamment de l'alphabet courant** → **9**. Par exemple, `\p{Ll}` trouve toute lettre minuscule alors que `\l` ne capture que les minuscules figurant dans l'alphabet. On a donc : `[a-z] ⊆ \l ⊆ \p{Ll}`
- De plus, l'interpréteur accepte un **sous-ensemble des raccourcis GREP** **B**.
- Tous les métacaractères recensés ici peuvent être utilisés seuls ou au sein d'une classe :
Ex. : `[\p{Ll}]`, `[~3~4~%]`, `[^\p{P}]`, etc.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de sortie



B
Fichier texte
Fichier XML
InDesign (extrait)

C
[À l'identique]
Automatique
Minuscules
Majuscules
Casse de titre
Majuscules > Auto

Chaque ligne représente une « entrée d'index ». Une tabulation (ou autre « marqueur d'entrée » → 26) sépare les sujets de la séquence de folios correspondants. Les membres éventuels (*Paris, Tostes, Yonville*) sont toujours regroupés et indentés sous leur sujet (*VILLES*).

D

```
27 soleil» 11, 17, 22, 27, 42, 48, 50, 58, 80, 82,
28 sommeil» 15, 194, 199
29 tristesse» 49, 51, 84, 122, 142, 167, 188, 346,
30 VILLES»
31 » Paris» 51, 62, 73-75, 77, 91, 101, 106, 109
32 » Tostes» 13, 15, 24, 39, 58, 70, 73, 76, 85-8
33 » Yonville» 89, 91-93, 100, 102, 106-109, 11
34 visage» 2, 45, 65, 77, 109, 120, 132, 142, 145,
```

23

PRODUIRE UN INDEX SOUS FORME DE FICHER TEXTE

La rubrique **Sortie** **A** rassemble les principaux paramètres d'export et de mise en forme de l'index.

1. Dans la liste **Destination**, choisissez « Fichier texte » (option par défaut) **B** pour recevoir l'index final dans un **fichier plain text**. Les autres options sont « Fichier XML » → 27 et « Extrait InDesign » (CS4/CS5) → 28.
2. Dans la liste **Casse**, choisissez « Automatique » (opt. par défaut) **C** pour laisser à IndexMatic² l'initiative de déterminer **la casse optimale** des entrées (l'algorithme sous-jacent est précisé → 24). Sauf mise en forme spécifique, cette option est la plus pertinente en général.
3. Dans la liste **Tri**, choisissez « [Latin générique] » (opt. par défaut) afin d'obtenir un classement correct des entrées d'index **D**. Il ne reste plus qu'à cliquer sur le bouton **Produire l'index**.

- Chaque fichier d'index est créé sous un nom unique **horodaté**, de la forme : *indexmatic_2011-07-01_20h52_21*. Ce fichier est placé dans le dossier contenant le document de travail. (Si le document n'a pas encore été enregistré, l'index est envoyé sur le Bureau.)
- La version **TRY** n'autorise que la destination « Fichier texte ».

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de sortie

bal	Bal	BAL
<i>bal</i>		
1	2	3

Le tableau ci-contre présente les résultats issus de différentes requêtes adressées au même document-test (ci-dessus).

- A** crée implicitement le TERME **bal** en minuscules. Le système n'a pas ici à tracer les variantes de casse, il produit un unique élément avec cette graphie.
- B** force le calcul du TERME (à travers **\$0**). Avec l'option « [À l'identique] », les variantes restent séparées. Avec l'option « Automatique », elles fusionnent sous la graphie la plus fréquente (*bal*).
- C** est sensible à la casse : la requête ne trouve donc que le vocable **Bal** en page 2. Le TERME implicite (**Bal**) est alors mis en forme selon l'option de casse.
- D** illustre la réécriture explicite et littérale d'une CLÉ (**bal**) dans un TERME (**Fête**).
- E** est similaire à **B**, mais le TERME calculé possède un sujet (**Fête**) et un membre variable **\$0**. Observez comment l'option « Maj. > Auto » agit sur ces éléments.



	REQUÊTE TESTÉE	bal/i	/ba\w/i->\$0	/Bal/I	bal/i=>Fête	/bal/i=>fête>\$0
CASSE		A	B	C	D	E
[À l'identique]	bal 1-3	Bal 2	BAL 2	Fête 1-3	fête bal 1 Bal 2 BAL 3	
Automatique	bal 1-3	bal 1-3	Bal 2	Fête 1-3	Fête bal 1-3	
Minuscules	bal 1-3	bal 1-3	bal 2	fête 1-3	fête bal 1-3	
Majuscules	BAL 1-3	BAL 1-3	BAL 2	FÊTE 1-3	FÊTE BAL 1-3	
Casse de titre	Bal 1-3	Bal 1-3	Bal 2	Fête 1-3	Fête Bal 1-3	
Maj. > Auto	bal 1-3	bal 1-3	Bal 2	Fête 1-3	FÊTE bal 1-3	

24

RÉGLAGE FIN DE LA CASSE EN SORTIE

Un TERME (→18) est stocké **provisoirement** dans un tampon mémorisant les **variantes de casse**. Cet aspect est imperceptible à l'utilisateur mais explique pourquoi IndexMatic² peut encore, si besoin, distinguer **animal** de **Animal**. Considérons la requête **/chats?/i**. Elle va rencontrer des formes comme **chat, Chats, CHAT, chats...** En interne, le script produit alors les modèles {chat} et {chats}, puis maintient un compteur pour chaque variante.

C'est ensuite qu'intervient le formatage des entrées selon l'option **Casse** : « [À l'identique] » exige que les variantes de casse soient reportées comme des **termes distincts**. « Automatique » fusionne les variantes et retient comme graphie unique la **casse la plus utilisée**. « Minuscules », « Majuscules » et « Casse de titre » ignorent les variantes et appliquent la casse correspondante. Enfin, l'option « Majuscules > Auto » compose les sujets en majuscules et les membres en casse « automatique » **si ces derniers sont présents**, sinon agit comme « Automatique ».

- Lorsqu'un **motif** ne contient **aucun** opérateur propre aux **regex**, IndexMatic² optimise silencieusement la requête en convertissant ce motif en simple **vocable**. Par exemple, **/chat/i** est converti en **chat/i** (sans barre oblique initiale). Ce faisant, le système produit un TERME implicite, **chat**, ignorant les **variantes de casse** (cf. **A**). Pour forcer le calcul des variantes, vous devez alors utiliser une variable : **/chat/i=>\$0**

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de sortie

Sortie

Destination : Fichier texte

Casse : Automatique

Tri : [Latin générique]

Sous-tri : Comme 1er niveau

[Pas de tri]
Comme 1er niveau
Séquence numérique

C

Retenez qu'IndexMatic² est profilé pour les langues basées sur l'alphabet latin → 9. Si vos termes d'index sollicitent des caractères extérieurs à cet alphabet — par exemple des caractères cyrilliques — le script ne saura pas les ordonner convenablement.

[Pas de tri]
[Latin générique]

Albanais
Basque
Bosniaque
Breton
Croate
Tchèque
Danois
Néerlandais (I=Y)
Néerlandais (moderne)
Anglais
Espéranto
Estonien
Îles Féroé
Filipino (moderne)
Finnois (multilingue)
Finnois (officiel)
Français
Frison
Allemand (umlauts décomposés)
Allemand (dictionnaire)
Groenlandais
Hongrois

Islandais
Indonésien
Irlandais (Gaélique)
Italien
Kurde (latinisé)
Letton
Lituanien
Luxembourgeois
Malais
Maltais
Moldave
Norvégien
Polonais
Portugais
Roumain
Serbe
Slovaque
Slovène
Espagnol (moderne)
Espagnol (traditionnel)
Suédois (moderne)
Suédois (V=W)
Turc
Turkmène
Gallois

EXEMPLE : Classement d'entrées lexicales issues d'un texte norvégien.

Tri: [Latin générique]	Tri: Norvégien
A	B
åbner	altså
ære	bryst
altså	brønder
brønder	brød
brød	fjord
bryst	jeg
fjord	ordet
jeg	vild
øje	ære
ordet	øje
vild	åbner

25

CLASSEMENT ALPHABÉTIQUE DES ENTRÉES D'INDEX

En bas du panneau « Sortie » figure un champ **Tri**, qui déploie dans une liste déroulante un grand nombre de choix. Chacun correspond à un **algorithme de classement alphabétique** propre à telle ou telle langue.

1. Choisissez **[Pas de tri]** pour désactiver complètement cette fonctionnalité. Les entrées d'index resteront non triées. Option utile si vous souhaitez préserver l'ordre original des TERMES fournis via l'Éditeur de requêtes.
2. Choisissez **[Latin générique]** **A** pour obtenir un ordre conforme aux règles communes de classement en Europe. Cette option est recommandée si votre ouvrage est multilingue ou contient des mots étrangers.
3. Choisissez un **algorithme spécifique** **B** pour appliquer les règles particulières à une langue-cible.
4. Les **membres** sont classés selon l'option de **Sous-tri** **C** : **[Pas de tri]**, **Comme 1er niveau** ou **Séq. numérique**.

■ Si l'option **[Pas de tri]** est sélectionnée durant une indexation en mode **automatique**, ou si des **termes variables** sont présents dans les requêtes, vous ne pouvez pas présumer du classement final des entrées, car IndexMatic² n'explore pas les documents de façon linéaire.

■ Si votre index contient des **membres** → 18, ces derniers sont ordonnés par défaut comme les sujets (Sous-tri > Comme 1er niveau). Choisissez « **Séquence numérique** » pour les ordonner numériquement.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de sortie

The screenshot shows the 'Présentation des pages' settings in IndexMatic 2.025. A red box highlights the 'Présentation des pages' section. Arrows point from specific settings to callout boxes labeled B, C, D, E, F, G, and H.

- A**: Points to the 'Présentation des pages' section header.
- B**: Points to the 'Marqueur d'entrée' field, which is set to 'Tabulation'. The callout box shows: Tabulation, Virgule + Tabulation, Deux-points + Tabulation, Espace, Virgule + Espace, Deux-points + Espace.
- C**: Points to the 'Séparateur de folios' field, which is set to '1, 2, 3'. The callout box shows: 1 2 3, 1, 2, 3, 1; 2; 3, 1 | 2; 3, 1 | 2 | 3, 1 • 2 • 3.
- D**: Points to the 'Regroupement' field, which is set to '10-20 (insécable)'. The callout box shows: [Pas de regroupement], 10-20 (insécable), 10-20 (demi-tiret), 10-20 (tiret long), 10_20 (tiret bas), 10/20 (barre oblique).
- E**: Points to the 'Empan' field, which is set to '2'.
- F**: Points to the 'Tolérance' field, which is set to '0'.
- G**: Points to the 'Ajout d'un point final' checkbox, which is unchecked.
- H**: Points to the 'Ignorer les éléments introuvables' checkbox, which is checked.

26

MISE EN FORME DES FOLIOS

La rubrique **Présentation des pages** **A** permet de structurer les **entrées** et les **folios** dans l'index final :

- 1. Marqueur d'entrée** **B** : chaîne à insérer entre une entrée et ses folios (par défaut, tabulation).
Séparateur de folios **C** : chaîne à insérer entre les folios (modèle par défaut : « **1, 2, 3...** »).
- 2. Regroupement** **D** : mise en forme des folios regroupés (s'il y a regroupement). Choisissez « [Pas de regroupement] » pour forcer le report individuel des pages. **Empan** (2-10) **E** : nombre minimum de **folios consécutifs** pour induire un regroupement. **Tolérance** (0-5) **F** : nombre de **folios manquant** (« trous ») autorisés dans un regroupement. Cochez **Ajout d'un point final** **G** pour conclure chaque ligne d'index par un point.
- 3. Ignorer les éléments introuvables** **H** : cochez cette case (activée par défaut) pour empêcher le report des **entrées non trouvées**. À l'inverse un **tiret cadratin** (—) servira d'indicateur : **PasTrouvé** —

■ Les **noms de pages** sont triés numériquement à condition que ce soit possible. Si le(s) document(s) cibles, ou certaines sections, utilisent des **styles de numérotation** spéciaux — tels que « i, ii, iii, iv... » ou « a, b, c... » — alors les pages correspondantes sont reportées sous leur nom propre **au tout début** de la séquence de folios. Exemple :
sujet vi, vii, xi, 53, 130-131...
(Ces folios spéciaux ne sont pas regroupés.)

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de sortie

Sortie
Destination : Fichier XML
Casse : Automatique

Indexation XML d'une mise en page de Hamlet (échantillon).

```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <index created="Wednesday, June 19 2011 19:35:17" scope="Hamlet.indd">
3   <topic hits="4">
4     <term>affair</term>
5     <locations pageRankCondition="3"></locations>
6   </topic>
7   <topic hits="4">
8     <term variants="Affection">affection</term>
9     <locations pageRankCondition="3"></locations>
10  </topic>
11  <topic hits="3">
12    <term>FAMILY</term>
13    <subtopic hits="16">
14      <term>brother</term>
15      <locations pageRankCondition="2">xi, 109, 183</locations>
16    </subtopic>
17    <subtopic hits="69">
18      <term>father</term>
19      <locations pageRankCondition="2">xii, 14-16, 18-19, 22, 35,
20    </subtopic>
21    <subtopic hits="44">
22      <term variants="Mother">mother</term>
23      <locations pageRankCondition="2">103, 109, 112, 122, 129</l
24    </subtopic>
25  </topic>
26  <topic hits="5">
27    <term>love</term>
28    <locations pageRankCondition="3">55, 95-97, 152</locations>
29  </topic>
30  <topic hits="22">
31    <term variants="Madness">madness</term>
32    <locations pageRankCondition="3">119</locations>
33  </topic>
34  <topic hits="16">
35    <term>madness</term>
```

topic (sujet)
(peut contenir l'attribut 'hits', si disponible)

locations (folios)
(la condition de page rank est rappelée dans un attribut)

subtopic (membre)
(même structure que 'topic')

En-tête
term (terme)
(un attribut 'variants' peut être renseigné si les variantes de casse sont analysées)

Notez que l'élément **locations** peut rester vide. En effet, la requête n'a pas été satisfaite (en raison d'un page rank élevé), mais les paramètres de sortie réclament ici le signalement des entrées « non trouvées ».

27 EXPORT XML

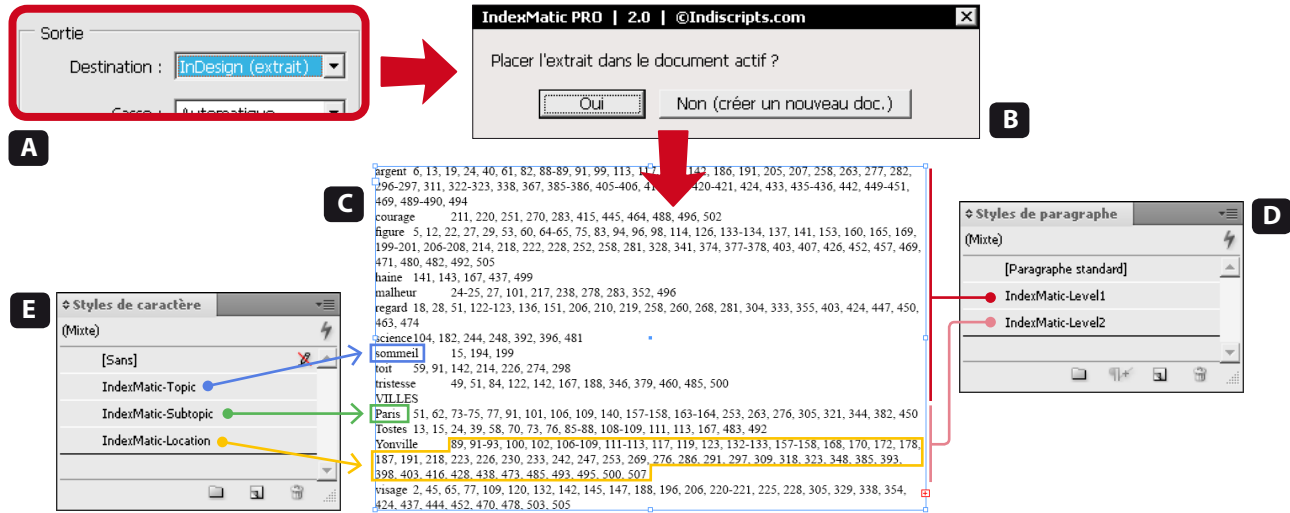
1. Sélectionnez la destination « **Fichier XML** » **A** dans la rubrique Sortie.
2. Ajustez à votre convenance les autres options : **Casse** → **24**, **Tri** → **25** et **Présentation des pages** → **26**. Ces paramètres ont une signification et un rôle inchangés dans le cadre d'un export XML.
3. Cliquez sur le bouton **Produire l'index** pour générer le fichier **B**. La structure XML résultante est **plus ou moins prolix** selon les paramètres de travail tels que le page rank, la casse de sortie ou autres réglages agissant sur le traitement des requêtes. Cependant, la structure minimale d'un élément est la suivante :
`<topic><term>...</term><locations>...</locations></topic>`

- L'export XML est désactivé dans la **version TRY**.
- Chaque fichier d'index est créé sous un nom unique **horodaté**, de la forme : `indexmatic_2011-07-01_20h52_21.xml`. Ce fichier XML est placé dans le dossier contenant le document de travail. (Si le document n'a pas encore été enregistré, l'index est envoyé sur le Bureau.)

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de sortie



28

PRODUIRE L'INDEX DANS INDESIGN (CS4/CS5)

1. Sélectionnez la destination « **InDesign (extrait)** » **A** dans la rubrique Sortie.
2. Ajustez à votre convenance les autres options : **Casse** → **24**, **Tri** → **25** et **Présentation des pages** → **26**. Ces paramètres ont une signification et un rôle inchangés dans le cadre d'un export vers InDesign.
3. Cliquez sur le bouton **Produire l'index**. Vous pouvez choisir de placer l'objet dans le document actif ou dans un nouveau document **B**. L'extrait InDesign est chargé dans le **curseur d'importation**. Il vous suffit alors de cliquer dans le document pour « réveiller » le bloc-texte sous-jacent.
4. Observez qu'aucun enrichissement particulier n'est appliqué aux entrées **C**. IndexMatic² choisit plutôt de créer des **styles de paragraphe D** et de **caractère neutres E** afin que vous puissiez les ajuster selon vos besoins.

- Cette fonctionnalité est désactivée dans la **version TRY**, et non implémentée dans **InDesign CS3**.
- À savoir : une fois paramétrés par vos soins les **styles IndexMatic** par défaut, ces derniers sont **préservés** lorsque vous régénérez et placez un nouvel index dans le **même** document de travail. Cela vous évite donc de « re-styler » constamment l'index.

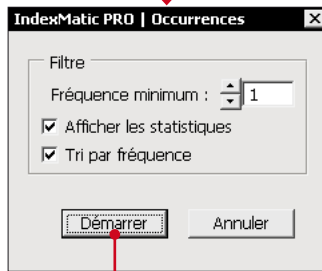
IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

Options de sortie



A



ITEMS	HITS		
1		38 grands	63
2		39 maintenant	62
3	Charles	40 Vonville	61
4	madame	41 revolt	61
5	Bovary	42 Souvent	61
6	quelque	43 disait	60
7	Homais	44 Comment	58
8	encore	45 Monsieur	57
9	contre	46 Pourquoi	56
10	Rodolphe	47 personne	56
11	toujours	48 figure	56
12	devant	49 milieue	54
13	autres	50 aurait	54
14	étaient	51 regardait	54
15	toutes	52 Fallait	54
16	cependant	53 pourtant	53
17	faisait	54 entendait	53
18	pharmacien	55 voyait	53
19	avait	56 grande	52
20	maison	57 dit-il	52
21	petite	58 lui-même	52
22	allait	59 enfant	52
23	heures	60 beaucoup	52
24	ailleurs	61 argent	51
25	autour	62 choses	51
26	jamais	63 répondait	51
27	chambre	64 pendant	51
28	est-ce	65 quatre	50
29	pauvre	66 eglise	50
30	venait	67 petits	49
31	tandis	68 lorsqu	48
32	semblait	69 doigts	48
33	Heureux	70 depuis	48
34	chaque	71 trouvait	48
35	derrrière	72 longue	47
36	quelques	73 ensuite	47
37	peut-être	74 francs	47
		75 restait	46
		76 pouvait	46
		77 félicité	46
		78 Femmes	46
		79 tenait	45
		80 médecin	45
		81 dit-elle	45
		82 demanda	45
		83 Fenêtre	45
		84 soleil	44
		85 autrefois	44
		86 lèvres	43
		87 sentait	43
		88 savait	43
		89 portait	43
		90 presque	42
		91 serait	42
		92 prendre	41
		93 aperçut	41
		94 prenait	41
		95 Rouault	40
		96 chercher	39
		97 papier	38
		98 passait	38
		99 jardin	37
		100 Justin	37
		101 bouche	37
		102 quelle	37
		103 lendemain	37
		104 cheval	37
		105 voulut	36
		106 campagne	36
		107 chapeau	35
		108 trouva	35
		109 travers	35
		110 devait	35
		111 poitrine	35

Les mots de *Madame Bovary* extraits par IndexMatic² en mode automatique →12.

29

OCURRENCES ET STATISTIQUES AVANT INDEXATION

Le bouton **Occurrences...** est apparu dans la version 2.025. Son objet principal est de produire les entrées **sans folios associés**. Usage : préparer une liste de mots-clés, tester des requêtes, établir des statistiques...

1. Une fois vos requêtes et options définies, cliquez sur **Occurrences...** pour ouvrir le dialogue afférent **A**.
2. La zone **Fréquence minimum** (1–100) permet d'exclure les termes n'apparaissant pas au moins *N* fois dans la portée →4. La valeur par défaut est 1 (filtre inactif). Augmentez la fréquence pour exclure les mots rares.
3. Cochez **Afficher les statistiques** pour que le rapport indique la fréquence de chaque terme. Si vous décochez cette case, la valeur n'est pas affichée mais cela n'inhibe pas le filtre Fréquence minimum.
4. **Tri par fréquence** range les entrées par fréquence décroissante **B** plutôt que selon les options de tri →25.

- Chaque rapport d'occurrences est créé sous un nom unique **horodaté**, de la forme : *indexmatic-hits_2011-10-04_20h52_21.txt*, placé dans le dossier du document de travail.
- Remarquez que le filtre **Fréquence minimum** agit **APRÈS** l'exécution des requêtes. Donc, si un terme est écarté en raison du **Page Rank** →10, il n'apparaîtra pas dans le rapport quelle que soit par ailleurs sa fréquence dans la portée indexée.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

EXTRACTION DU CONTENU DES PAGES

Bien qu'IndexMatic adapte ses méthodes d'investigation à vos réglages (portée, contexte, styles, options de concordance...), le script doit explorer jusqu'à un certain point la structure intime des documents pour identifier le « contenu » de chaque page. Sous InDesign, les blocs chaînés, objets ancrés, tableaux et notes peuvent sensiblement compliquer cette opération. Voici quelques précisions sur les procédures adoptées, et quelques conséquences pratiques.

1. CHAÎNAGES ET SAUTS

IndexMatic commence par recenser les blocs-texte rencontrés sur les pages de la portée spécifiée. Un bloc-texte, simple conteneur visuel, reflète en partie des flots textuels — article, note(s), cellule(s) de tableaux — qui se développent chacun de façon autonome. Ainsi, une simple page peut adresser plusieurs bribes de plusieurs flots. Il arrive souvent qu'un flot se poursuive d'une page à l'autre, ou se distribue dans plusieurs blocs d'une même page, ou encore qu'il s'interrompe brutalement (texte en excès).

Le processus d'indexation consiste à extraire ce qui apparaît, pour le lecteur, comme la continuité textuelle d'une page donnée. Par exemple, lorsqu'un mot est coupé entre deux blocs chaînés sous l'effet d'une césure, l'indexeur doit ignorer cette rupture et considérer que le mot apparaît complet, même si un fragment est rejetée sur une autre page. Toutefois, si deux blocs-texte se jouxent (sur la même page) mais correspondent à des articles distincts, l'indexeur ne doit pas évidemment pas les regarder comme chaînés.

2. PRÉ-LECTURE

Supposons que nous souhaitions indexer les occurrences de l'expression « Alan Turing » au sein d'un document constitué de blocs chaînés. Quand l'expression recherchée est à *cheval* sur deux pages (« Alan » en page A, « Turing » en page B), IndexMatic doit déceler qu'une occurrence complète de « Alan Turing » existe sur la page A. Pour ce faire, le script procède à une pré-lecture au-delà de la page courante et gère un tampon de données portant sur la page à venir. Pour des raisons de performance, **la taille de ce**

tampon est limitée à 16 caractères. Par conséquent, IndexMatic ne peut pas détecter certaines concordances « longues » franchissant l'inter-pages.

3. NOTES DE BAS DE PAGE ET TABLEUX

Avant la version 2.025, IndexMatic rattachait **la totalité d'une note** à la page d'appel initial, même si ladite note ne tenait pas sur une seule page du fait de sa taille. Ce point a été corrigé : les « suites » d'une note de bas de page sont désormais correctement localisées.

De même, le script localise correctement les **cellules d'un tableau** distribué sur plusieurs pages. Toutefois, IndexMatic continue d'ignorer **les tableaux de second niveau**. Notez également que la routine d'analyse des cellules est indépendante de l'algorithme principal et peut accroître sensiblement le temps d'exécution.

4. PARAGRAPHES ET RÉGIONS DE STYLES

Chaque fois qu'IndexMatic identifie une portion textuelle pertinente sur une page donnée (partie d'un article, d'une note, ou cellule), il stocke l'information dans une structure spécifique appelée « train de page » (*page run*). **Chaque train de page est alors divisé en « wagons » (segments) qui correspondent aux plus longues unités de texte qu'une requête puisse adresser.** Aucun wagon ne peut excéder la taille d'un paragraphe. C'est pourquoi une requête maximale comme **/./+/e** ne peut pas embrasser davantage qu'un paragraphe (retour-chariot exclu).

Certains wagons sont même plus petits qu'un paragraphe. Par exemple, lorsqu'un style de caractère est ciblé, les wagons coïncident avec les seules régions soumises à ce style et, dans ce cas, une requête ne peut pas « traverser » plusieurs régions. Cependant, lorsqu'un **groupe de styles de caractères** est ciblé, les régions adjacentes relevant de ce groupe sont fusionnées en un seul wagon, ce qui permet typiquement d'adresser des régions de styles hétérogènes, du moment qu'elles reflètent bien le groupe-cible.

IndexMatic² Manuel

© Indiscripts.com | Octobre 2011 | v. 2.025

LIMITATIONS, PROBLÈMES CONNUS

1. LANGUES

La version actuelle d'IndexMatic ne gère pas les écritures droite-à-gauche (ou les textes bi-directionnels). Le script est conçu pour explorer des documents basés sur l'alphabet latin. Cependant, tout caractère Unicode peut être recherché par une requête *via* la syntaxe `\uHHHH`.

2. TRAITEMENT DES LIVRES

Lors du traitement d'un livre InDesign, IndexMatic ouvre en tâche de fond les documents sous-jacents pour procéder à certains contrôles et préparer l'indexation. Ce mécanisme est normalement transparent pour l'utilisateur et ne cause pas d'erreur. Toutefois, certains beta-testeurs ont signalé un bug mal élucidé concernant la gestion des livres. Il semble que dans certains cas IndexMatic échoue à analyser le statut interne d'un chapitre ou détecte une « collision » entre le nom d'un chapitre et le nom d'un autre document ouvert. Ce bug concernerait plus particulièrement les livres ou chapitres réimportés depuis une version d'InDesign plus ancienne. Auquel cas, nous vous suggérons de reconstruire le livre pour voir si le problème persiste. En dernier recours, veuillez ouvrir manuellement tous les chapitres à indexer avant d'exécuter le script.

3. INTERPRÉTEUR DE REQUÊTES

- (a) IndexMatic n'est pas basé sur le module Grep d'InDesign ; en conséquence il ignore certains opérateurs strictement rattachés à Grep ainsi que les raccourcis Posix. Référez-vous aux sections **OPÉRATEURS SPÉCIAUX DANS LES REGEX** et **MÉTACARACTÈRES ADDITIONNELS** pour une vue exhaustive des métacaractères implémentés.
- (b) La CLÉ d'une requête est limitée à 172 caractères.

VERSIONS

2.025 8 octobre 2011

MÀJ MAJEURE. Correction d'un bug lié à l'utilisation de classes complémentaires `[^xxx]`. Correction d'un problème de « segmentation » du texte au niveau des blocs chaînés et des régions repérées par des styles de caractère. Localisation correcte des notes de bas de page multi-pages. Amélioration de l'algorithme d'extraction des contenus inter-pages. L'espace générique reconnaît dorénavant le saut de ligne forcé. Taille limite de clé étendue à 172 caractères. Ajout des métacaractères : `\w, \1, \L, \m, \M, \p{...}`. Ajout du « Marqueur d'entrée ». Prise en charge des commentaires et de la syntaxe de renvoi (`// sujet1 => V. sujet2`). Ajout de l'option « Inclure le tiret bas » (alphabet). Nouvelle interface de l'Éditeur de requêtes, boutons ouvrir / sauvegarder, augmenter / réduire taille des caractères. Ajout de la fonction « Occurrences ».

2.024 23 juillet 2011

MÀJ IMPORTANTE. Correction de certains problèmes liés à l'appui de la touche Entrée dans l'Éditeur de requêtes (Mac). Résolution d'un bug sérieux portant sur la capture des mots en mode Automatique et sur l'emploi du métacaractère `\w` dans les expressions régulières. Ajout de la fonctionnalité de « Sous-tri » applicable aux entrées d'index de second niveau.

2.022 6 juillet 2011

MÀJ MINEURE. Ajustement de l'algorithme traitant l'option « Espace générique ». Correction de coquilles affectant les bulles d'aide de la boîte de dialogue principale.

2.021 1^{er} juillet 2011

VERSION PUBLIQUE ORIGINALE.



ALPHABET 8-9, 11-13, 21-22, 25, 31
→ aussi **CARACTÈRES**

langue 25, 31

ANCRÉS (OBJETS ~) 6, 21, 30

BLOC-TEXTE 4, 28, 30

article 4, 30

~ chaîné 30-31

~ en excès 30

CALQUE 4-5

CAPTURE → **CONCORDANCE**

CARACTÈRES

apostrophe 9, 22

Ascii 9

casse des ~ → **CASSE**

chiffres 9, 17, 21-22

~ conditionnels 9, 21

~ d'espace → **ESPACE**

diacritiques 9

~ insécables 9, 11, 22

lettres 9, 21-22

méta ~ → **REGEX**

symboles 18-19, 21-22

tabulation 11, 21-23, 26

tiret 9, 22, 26, 31

CASSE 11, 16-17, 19, 21-24, 27-28

CONCORDANCE 9, 11-13, 17, 19, 30

capture 3, 9, 15, 19, 21-22, 31

formes, variantes 8, 11, 15, 19, 24, 27

variables → **REQUÊTE**

CONTEXTE 6, 8, 20, 30

EMPAN → **PAGES**

ESPACE 11-13, 15-17, 21-22, 31

~ générique 11, 16-17, 21, 23, 25, 31

EXPRESSIONS RÉGULIÈRES → **REGEX**

FILTRE 6-7, 12, 29

FOLIOS → **PAGES** / *folios*

INDEX 1-10, 12-15, 18-31

entrée d' ~ 14, 23-29, 31

export 23, 27-28

mise en forme 11, 23-24

pages et folios → **PAGES**

séparateur, marqueur 20-23, 26, 31

tri et classement 23, 25-29, 31

LIVRE (INDESIGN) 2, 4-5, 7, 31

MÉTACARACTÈRES → **REGEX**

NOTE (DE BAS DE PAGE) 6, 30-31

OCCURRENCES 1-3, 10-11, 21, 27, 29-31

PAGE RANK 2-3, 10, 12-13, 17, 27, 29

PAGES 2-4, 6, 8, 10, 12-13, 17, 21, 24, 26-31

empan 26

folios 2-4, 20, 23, 26-27, 29

localisation 27

regroupement 15, 26

tolérance 26

PARAGRAPHE 7, 21, 28, 30

PARENTHÈSES → **REGEX**

PORTÉE 3-6, 8-9, 29-30

RÉÉCRITURE → **REQUÊTE** | *opérateurs*

RÉFÉRENCES CROISÉES → **REQUÊTE** | *renvois*

REGEX 2-3, 8-9, 12, 16, 18-19, 21-22, 24, 31

classe (de caractères) 21-22, 31

concordance → **CONCORDANCE**

Grep 2, 11, 22, 31

métacaractères 9, 16, 21-22, 31

motif 9, 14, 16, 18-19, 21, 24

parenthèses 19, 21

point 11, 21-22, 26, 30

Unicode 9, 11, 21-22, 31

→ **REQUÊTE**

RENOVOIS → **REQUÊTE**

REQUÊTE 2, 8-10, 12-22, 24-25, 27, 29-31

clé 16-21, 24, 31

commentaire 14, 20, 31

flags 17

interpréteur 15, 18, 20, 22, 31

motif → **REGEX**

opérateurs 2, 15-16, 18-21, 24, 31

pseudo-terme 20

renvois 14, 20, 31

sujet / membre 14-15, 18-20, 23-27, 31

terme 1-3, 12-13, 15, 18-21, 24-25, 27, 29

variables 19-21, 24-25

vocabulaire 14, 16, 18-19, 24

SOUS-TRI → **INDEX** | *tri et classement*

STATISTIQUES 29

TABLEAUX ET CELLULES 6, 21, 24, 30

TABULATION → **CARACTÈRES**

TIRET → **CARACTÈRES**

TOLÉRANCE → **PAGES**

TRI → **INDEX** | *tri et classement*

XML 2, 23, 27