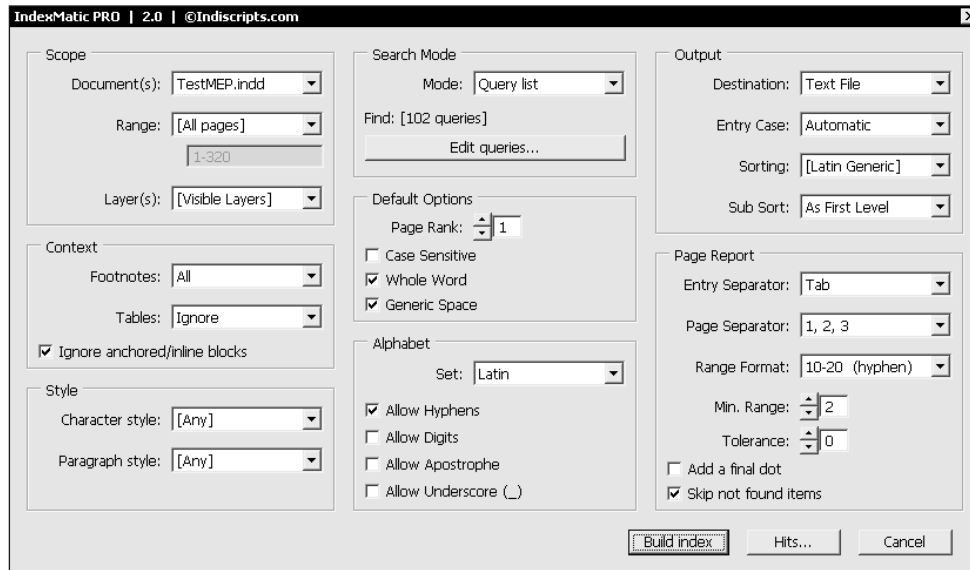


# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Getting started



Main dialog of IndexMatic<sup>2</sup> (2.025, Windows UI).

## 1 INSTALLING & LAUNCHING INDEXMATIC<sup>2</sup>

1. Download IndexMatic<sup>2</sup> from <indiscripts.com>. [PRO Version] Retrieve the package from your private link and save it on your hard disk. [TRY Version] Right-click **IndexMaticTry.zip** and choose Save Target / Link As.
2. Extract **IndexMaticPro.jsx** (or **...Try.jsx**) from the zip file, then place the file into your InDesign scripts folder (usually: **Scripts Panel**).
3. Start InDesign, open a document, open the Scripts panel: **Window > Utilities > Scripts** (CS5), or **Window > Automation > Scripts** (CS3, CS4). Then double-click on **IndexMatic...jsx**.

- IndexMatic<sup>2</sup> supports **InDesign CS3, CS4 or CS5+** (recommended) and works on both **Mac and Win** platforms.
- The **TRY version** has some limitations: you cannot change the output format (Destination: **Text File**), the final index is limited to **50 terms**, you have not access to the **Hit report**, and the script does not backup the **current settings**.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Getting started

**A**

**B**

**C**

1 analysis	9-10, 23, 33, 57-58, 61, 65, 75
2 anxiety	3, 5, 9-11, 13-14, 16, 18-21, 23
3 choice	33, 36-44, 47, 49, 51-53, 73, 114
4 conflict	14-18, 20-21, 25, 73, 146, 189
5 desire	17-18, 20, 24-25, 30, 38-39, 42
6 discourse	10-11, 23, 26, 57, 65, 71-72, 79
7 dream	38, 61, 74, 80, 85, 91, 93, 137
8 existence	34, 69, 107-108, 110, 114, 118-119
9 FAMILY	44, 52, 59, 76, 90-91, 122-125, 127
10 child	38, 41, 46, 78, 87-88, 91-93, 94
11 children	27-28, 41, 46, 94, 122-124, 134
12 father	38, 65, 87-89, 100-104, 106, 108
13 mother	41-42, 60, 78, 87-93, 96, 122-124
14 fantasy	19-20, 44, 56, 77, 115-116, 119
15 health	44, 47-51, 123, 199
16 ideology	33, 35-37, 44-49, 195
17 metaphor	58, 89, 96, 150-151, 157, 192
18 mirror	41
19 paradox	17, 19, 49, 85, 101, 114, 118
20 school	7, 27-28, 131, 135-136, 138, 189
21 society	15, 33, 35, 40, 44, 50, 52-53, 71
22 suicide	103, 110, 145, 147-148, 151-154
23 traumatism	10, 19, 25-26, 28, 40, 42-44, 56
24 virtual	26-28, 30, 52

## 2 QUICK OVERVIEW

1. IndexMatic<sup>2</sup> helps you to create an independent **index** from any InDesign document or book **A**. You can use it to automatically build subject, language, or author indexes. The script provides an extended set of methods that inspect the target document(s), track every occurrence of the desired keywords or expressions, and report the corresponding **page numbers**.
2. IndexMatic<sup>2</sup> is based on a sophisticated **query engine** **B** that allows many refinements through regular expressions, style filtering, 'page rank', topic rewriting...
3. The PRO version offers several output formats: **plain text** file **C**, **XML** file, InDesign **snippet** (CS4/CS5).

- IndexMatic<sup>2</sup> sidelines the native InDesign 'Index' feature. The script **never alter existing document(s)**: it just 'scans' the contents.
- Compared to other search tools, IndexMatic<sup>2</sup> usually provides excellent performance in that it **does not** employ the **GREP** engine at all.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Getting started

**A** Scope  
Document(s): Active Document (t)  
What document(s) do you need to process?  
Range: [All pages]  
1-204  
Layer(s): [Visible Layers]

**B** Search Mode  
Mode: Automatic  
Min. Length: 6  
Max. Length: 10  
Maximum number of characters per string.  
Page Rank: 1

**C** Default Options  
Page Rank: 3  
The Page Rank, from 1 to 9, represents the minimum number of times an expression must occur on the same page to be indexed at the corresponding page number. A specific query can explicitly override the default Page Rank, using the syntax 'xxx/1', 'xxx/2', etc.  
Alphabet

**D** E

```
187 millennium 33
188 mirror 41
189 missing 30
190 modern 34
191 moment 63, 92, 108, 111, 180
192 months 123
193 mother 78, 88, 91-92, 122-125, 129, 135-137, 14
194 Mourning 167
195 movement 108
196 nature 15
197 negative 171
198 neurosis 42, 133-134
199 neurotic 118, 133
200 nightmare 189
201 normal 88
202 Norton 197
203 object 18-19, 42, 75, 78, 89, 92-93, 101-102, :
204 objects 181
205 Oedipus 87, 168
206 oneself 48
207 opaque 117
208 others 39
209 Oxford 70
210 painful 94
211 paradox 85
212 parents 38, 121-123
213 particular 42
214 partner 51, 116
215 passage 113, 117, 164-165
216 patients 50
217 people 33, 37, 48, 50-52
218 perceived 36, 49
219 perception 114
220 perhaps 61
221 person 43, 53
```

## 3 INDEXING YOUR FIRST DOCUMENT IN 15 SECONDS

The simplest way to test IndexMatic<sup>2</sup> is to use the **Automatic** search mode (default). This hides the details of the query system and allows to quickly produce an index from scratch, without any explicit set of queries.

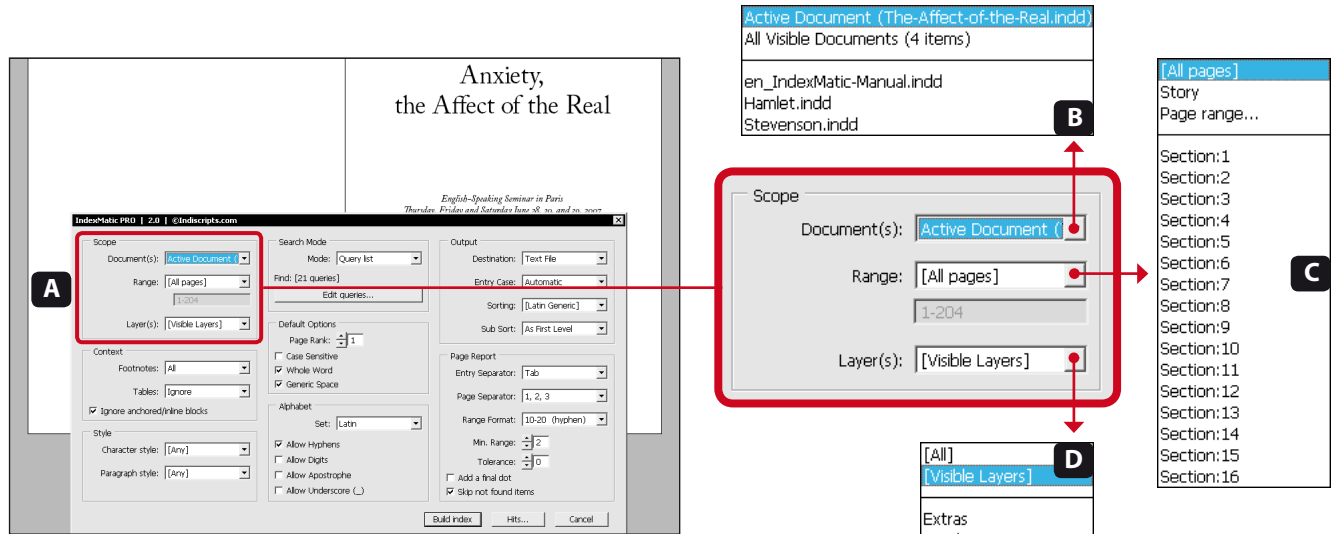
1. Open an InDesign document and launch the script. Make sure '**Active Document**' is highlighted as the scoped document (**Scope** panel) **A**
2. In the **Search Mode** panel, select '**Automatic**' **B** and adjust the **Min.** and **Max. Length** **C** of the words that you want to capture.
3. In the **Default Options** panel, set the **Page Rank** to 3 or 4 **D**. Finally, click **Build index** to produce the file **E**.

- Behind the scene, the **Automatic** search method formats and sends a series of **regular expressions (regex)** to the main module →15.
- If you need to extract word stats without page numbers, use the special **Hits...** button →29. (This feature is only available in the PRO version.)

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025

Scope & Context



## 4 THE SCOPE PANEL

The **Scope** panel **A** indicates which document(s) to explore and defines the work area for indexing.

1. The **Document(s)** field provides a dropdown control which lists every available document **B**. The active document is highlighted by default. You can choose a different one, or even target all open documents.
2. When a single document is targeted, the **Range** list **C** allows you to restrict the pages on which to operate. You can directly enter a **page range**—e.g. **10-20**; **25**; **30-40**—or select a specific **section** of the document. If a text frame is selected in InDesign, you can also focus on the related **Story**.
3. In addition, the **Layer(s)** list **D** gives you the option to process a specific layer, or only the **visible** ones.

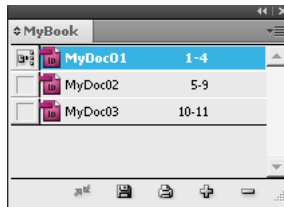
- When the scope includes **All Visible Documents** no distinction is made between **identical page numbers** that might belong to different documents. It is your responsibility to set the proper **page numbering** for each indexed document. (*About book indexing* → 5)

# IndexMatic<sup>2</sup> Manual

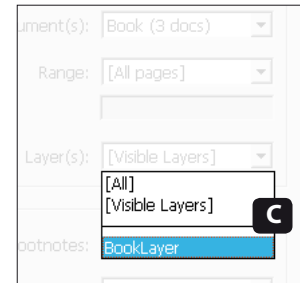
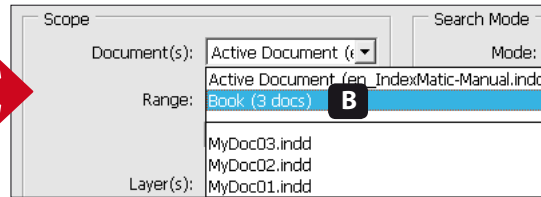
© Indiscripts.com | October 2011 | v. 2.025



Scope & Context



A



## 5 INDEXING A BOOK

Indexing an **entire book** with IndexMatic<sup>2</sup> is almost as easy as indexing a particular document:

1. Open the book file in InDesign **A**. (You don't need to open the underlying documents.)
2. Launch IndexMatic<sup>2</sup> and highlight the **'Book'** item in the **Document(s)** dropdown list **B**. The list item indicates the number of associated documents. (Note that each book's document also appears separately in the list, which gives you the option to process a specific chapter.)
3. If needed, select the proper layer in the **Layer(s)** list **C**. When you are targeting a book, the list only displays the **layer names** that belong to **every** document (i.e. the 'common layers'.)

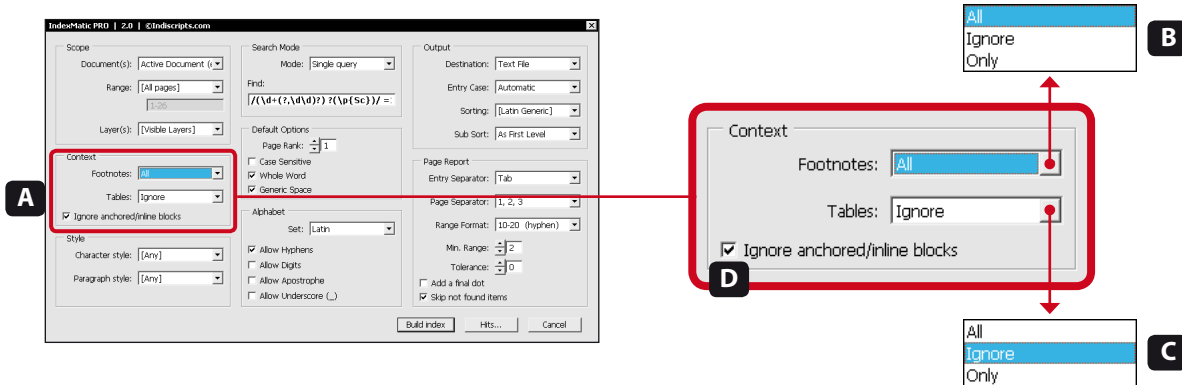
- IndexMatic<sup>2</sup> cannot manage several books at the same time. When indexing a book, make sure that a **single book file**, i.e. a single tab, appears in the Book panel.
- To index **only a few chapters**—not the entire book,—you must open (double-click) each corresponding document in InDesign and to target **All Visible Documents**.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Scope & Context



## 6 REFINING THE 'CONTEXT'

The **Context** panel **A** provides three additional filters over the Scope settings:

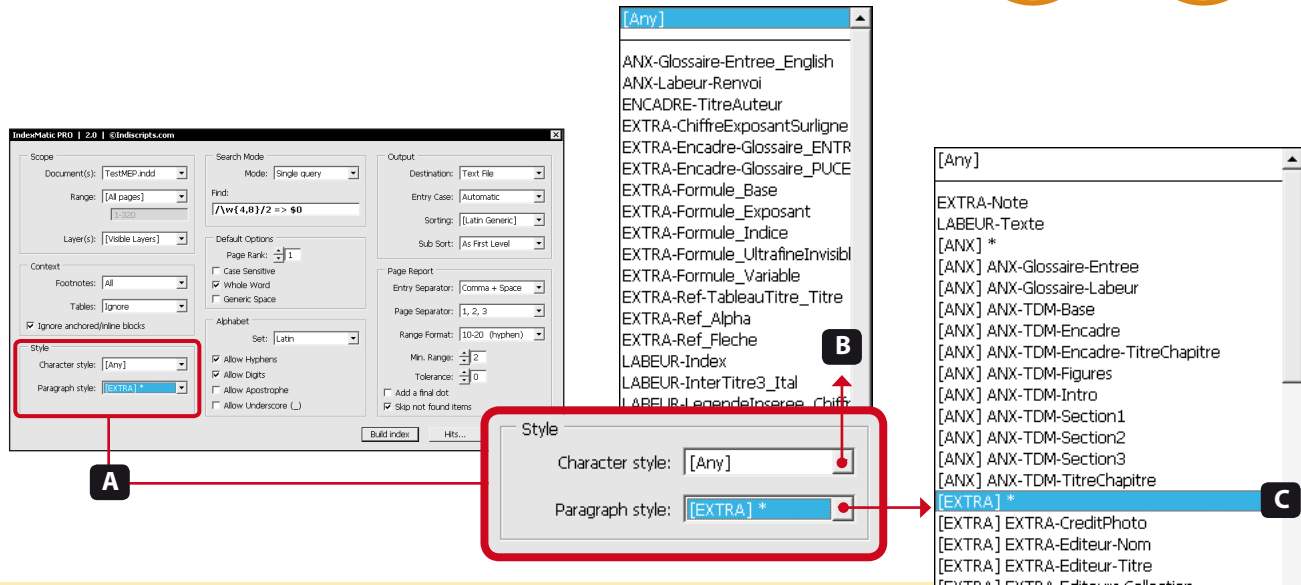
1. The **Footnotes** field allows to extend or restrict the search area to footnotes: in the dropdown list **B**, choose 'All' to include the footnote contents (default), choose 'Ignore' to disregard footnotes, and choose 'Only' to specifically restrict the search area to footnotes (within the current scope).
2. The same logic applies to the **Tables** filter **C**, which extends or restrict the search area to InDesign tables within the current scope. Note that IndexMatic<sup>2</sup> can **scan only first-level cells**: nested tables are ignored.
3. Check **Ignore anchored/inline blocks** **D** to disregard the contents of any embedded text frame (default).

- Since footnotes can contain tables, combining **Footnotes:Only** and **Tables:Only** makes sense, but this is very restrictive!
- How you set the **Context** filters may significantly impact the execution time.
- Since the version 2.025, issues related to **footnotes** and/or **tables** indexing have been solved.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025

Scope & Context



## 7 APPLYING PARAGRAPH/CHARACTER STYLE FILTERS

Well-structured documents use paragraph and character styles to give the designer maximum control over the layout. Styles are also useful to 'markup' the semantic pieces of your document: titles, headlines, captions, main text, product description, etc. Thanks to the **Style** panel **A**, you can tell IndexMatic<sup>2</sup> to selectively explore style-tagged contents within the current scope:

1. **Character style** **B** allows you to restrict the search field to a specific **character** style or style group.
2. **Paragraph style** **C** allows you to restrict the search field to a specific **paragraph** style or style group.

Style groups are listed in the form: **[group\_name] \*** (the star here means: "any style in that group.")

- Filtering contents by style is also available when you target a **book** or **multiple documents**. In this case the style lists only display **common** items.
- The search engine looks for applied **style(s)** regardless of the actual format of the text—hence local overrides have no effect on the parser.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Search Options

Search Mode  
Mode: **Automatic**

Min. Length:

Max. Length:

**A**

Search Mode  
Mode: **Query list**

Find: [21 queries]

Edit queries...

**B**

Search Mode  
Mode: **Single query**

Find:

**C**

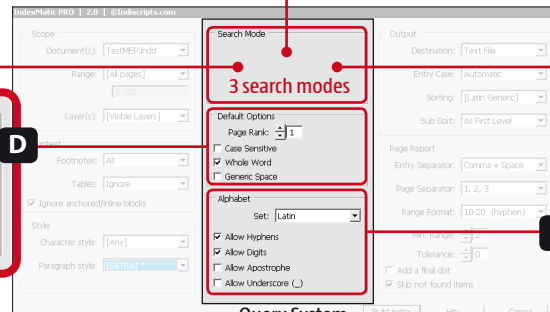
Default Options  
Page Rank:

Case Sensitive

Whole Word

Generic Space

**D**



Query System

Alphabet  
Set: **Latin**

Allow Hyphens

Allow Digits

Allow Apostrophe

Allow Underscore ( \_ )

**E**

# 8

## ABOUT THE QUERY SYSTEM

Once you have defined which parts of the document(s) you want to inspect (*see previous pages*), you'll tell IndexMatic<sup>2</sup> how to extract relevant words, in which form and under which conditions.

1. The **Search Mode** panel has three alternative states: **Automatic**, **Query List**, and **Single Query**. Choose **Automatic** **A** to find every word having a bounded number of letters. Choose **Query List** **B** to process a set of specific queries (word list, regular expressions...). Choose **Single Query** **C** to quickly test a query.
2. The **Default Options** **D** and **Alphabet** **E** panels show the **global settings** of the query system. These settings apply to all the search modes and specify how queries must be interpreted.

- During your InDesign session, IndexMatic<sup>2</sup> PRO stores all the settings made from the dialog box (scope, context, styles, queries, etc.), so you can easily refine the current parameters from the previous state.





# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Search Options

Eribusandiat omni autatendit quunt fuga. <b>Peirce</b> rit porersped quod este-cuptis dolupatatem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem volessitae premped utem con	Ucipicient vendis dolorehenis as <b>Peirce</b> aut offic tem conse-quam rehendit aut quiaue aut expe pel-liam rent carum sant <b>Peirce</b> excec idu-citissam idi dolorio qui saepa eiu? Cil-igent. Eprepre ven-
20	21

Eribusandiat omni autatendit quunt fuga. <b>Peirce</b> rit porersped quod este-cuptis dolupatatem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem volessitae premped utem con	Ucipicient vendis dolorehenis as <b>Peirce</b> aut offic tem conse-quam rehendit aut quiaue aut expe pel-liam rent carum sant <b>Peirce</b> excec idu-citissam idi dolorio qui saepa eiu? Cil-igent. Eprepre ven-
20	21

Eribusandiat omni autatendit quunt fuga. <b>Peirce</b> rit porersped quod este-cuptis dolupatatem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem volessitae premped utem con	Ucipicient vendis dolorehenis as <b>Peirce</b> aut offic tem conse-quam rehendit aut quiaue aut expe pel-liam rent carum sant <b>Peirce</b> excec idu-citissam idi dolorio qui saepa eiu? Cil-igent. Eprepre ven-
20	21

Eribusandiat omni autatendit quunt fuga. <b>Peirce</b> rit porersped quod este-cuptis dolupatatem re od ea dita sitio blab id ut ut aperiatem eatium facepro tem volessitae premped utem con	Ucipicient vendis dolorehenis as <b>Peirce</b> aut offic tem conse-quam rehendit aut quiaue aut expe pel-liam rent carum sant <b>Peirce</b> excec idu-citissam idi dolorio qui saepa eiu? Cil-igent. Eprepre ven-
20	21

impo iciunt et etur mo blaut et is ex et harupti quam illat-estiae niae alit mod mos et ex eum none vellatincius moles-to voluptaspis cum et et, ut mo volecti apeditam quia nese-quaehentisquisit	<b>Peirce</b> sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni- <b>Peirce</b> dus nos unti <b>Peirce</b> corem rendaer non-sequid enis utem in estis <b>Peirce</b> eiu?, cor-eror <b>Peirce</b>
22	23

impo iciunt et etur mo blaut et is ex et harupti quam illat-estiae niae alit mod mos et ex eum none vellatincius moles-to voluptaspis cum et et, ut mo volecti apeditam quia nese-quaehentisquisit	<b>Peirce</b> sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni- <b>Peirce</b> dus nos unti <b>Peirce</b> corem rendaer non-sequid enis utem in estis <b>Peirce</b> eiu?, cor-eror <b>Peirce</b>
22	23

impo iciunt et etur mo blaut et is ex et harupti quam illat-estiae niae alit mod mos et ex eum none vellatincius moles-to voluptaspis cum et et, ut mo volecti apeditam quia nese-quaehentisquisit	<b>Peirce</b> sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni- <b>Peirce</b> dus nos unti <b>Peirce</b> corem rendaer non-sequid enis utem in estis <b>Peirce</b> eiu?, cor-eror <b>Peirce</b>
22	23

impo iciunt et etur mo blaut et is ex et harupti quam illat-estiae niae alit mod mos et ex eum none vellatincius moles-to voluptaspis cum et et, ut mo volecti apeditam quia nese-quaehentisquisit	<b>Peirce</b> sum quos el-laccus mo volupta quia nonsedia quam eius con nonem. Quaeeti ameni- <b>Peirce</b> dus nos unti <b>Peirce</b> corem rendaer non-sequid enis utem in estis <b>Peirce</b> eiu?, cor-eror <b>Peirce</b>
22	23

Page Rank=1

Peirce 20, 21, 23

Page Rank=2

Peirce 21, 23

Page Rank=3

Peirce 23

Page Rank>3

Peirce —  
(not found)

## 10

### UNDERSTANDING THE 'PAGE RANK'

In IndexMatic<sup>2</sup>, a **page rank** is an integer—between 1 and 9—which refers to the **minimum number of hits** of a given expression in a given page. The underlying principle may be expressed intuitively as “the more a word occurs in a page the higher is the relevance of indexing that word.”

1. By setting a page rank greater than 1—say 3—in the **Default Options** panel, you force the query system to ignore any match that does not occur at least 3 times in a given page.
2. Increasing the page rank is particularly effective when using the **Automatic** search mode as it usually produces a large number of lexical units.

■ All **search modes** support the ‘page rank’ threshold, which acts globally. However, a singular query can **inhibit or override** the default page rank.

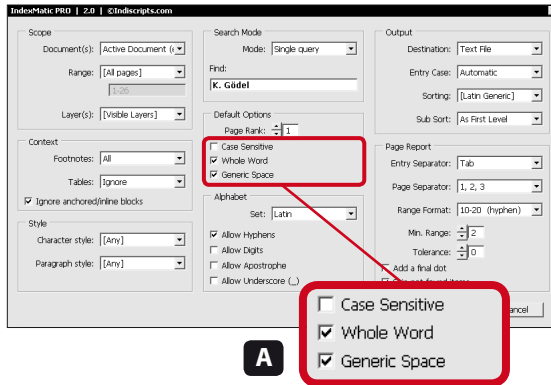
■ If a given expression occurs within the scope but never passes the ‘page rank’ test, then it is considered **not found** and can be reported as such →26.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Search Options



## B Generic Space Matching Table

The table lists every character that a simple space recognizes when Generic Space is enabled.

\*The **Forced Line Break** has been added in v. 2.025

InDesign Name	GREP	Unicode	Symb.
Space		U+0020	·
Forced Line Break*	\n	U+000A	↵
Nonbreaking Space (justif.)	~S	U+00A0	^
Nonbreaking Space (fixed width)	~S	U+202F	^
Tab	\t	U+0009	»
Indent To Here	~i	U+0007	↑
Right Indent Tab	~y	U+0008	↵
Flush Space	~f	U+2001	~
En Space	~>	U+2002	~
Em Space	~m	U+2003	~
Third Space	~3	U+2004	·
Quarter Space	~4	U+2005	·
Sixth Space	~%	U+2006	·
Figure Space	~/	U+2007	#
Punctuation Space	~.	U+2008	!
Thin Space	~<	U+2009	·
Hair Space	~	U+200A	·
Medium Mathematical Space		U+205F	·

## 11 MATCHING OPTIONS

The **Default Options** panel offers three options **A** that can either enforce or relax the matching constraints. Suppose we need to index the occurrences of **K. Gödel** (this could easily be done through a Single query):

1. By turning **Case Sensitive** on, we tell the search engine to exactly match the capitalization of the string, excluding variants such as **K. GÖDEL**. (This is usually *not* what we want, so the box is unchecked by default.)
2. The **Whole Word** option (checked by default) means that the string cannot be part of a larger word. More precisely, this means that **no alphabet's character** (→9) should precede or follow the match.
3. The **Generic Space** option (checked by default, when relevant) means that the space character between **K.** and **Gödel** may substitute **any InDesign white space** (non-breaking spaces, tabulations... **B**).

- Case Sensitivity regards characters (uppercase vs. lowercase) and **disregards text formatting**. For example, even if 'small cap' is applied to the letter 'a', the search engine still sees that character as a lowercase letter.
- Enabling **Whole Word** does not disallow the match to contain itself **inner spaces or extra characters**.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025

Search Options

Search Mode  
Mode: Automatic **A**  
Min. Length: 4 **B**  
Max. Length: 10 **B**

Default Options  
Page Rank: 4 **C**  
 Case Sensitive  
 Whole Word

Style  
Character style: [Any]  
Paragraph style: Main-Text **D**



1	Alexander	166
2	beauty	83
3	body	126
4	cannot	105
5	Captain	131
6	carriages	179
7	Clown	156-158, 162-164
8	come	182
9	death	138
10	doth	97
11	doubt	55
12	drink	188
13	duty	12
14	England	129
15	father	14, 18
16	First	75, 156-158, 162-164
17	Ghost	34
18	good	17, 120, 137
19	Hamlet	124, 182-183
20	hath	159
21	have	11, 64, 74, 77, 85,
22	heaven	111
23	help	112

## 12 USING THE AUTOMATIC MODE

The **Automatic** mode can typically be used to retrieve the **vocabulary** of a book, or to collect all **product names** from a catalog provided that paragraph and/or character styles are relevantly applied to those names.

1. In the **Search Mode** dropdown list, select 'Automatic' **A**.
2. Adjust the **Min. Length** ( $\geq 2$ ) and **Max. Length** ( $\leq 40$ ) **B**. The query system will search in the scope any sequence of characters forming a 'whole word' according to the rules of the current **alphabet**  $\rightarrow$  **9**.
3. If you are extracting a vocabulary, consider to increase the **page rank** **C** in order to retrieve the most significant occurrences. Also, apply a **style filter** **D** when necessary  $\rightarrow$  **7**.

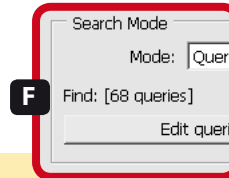
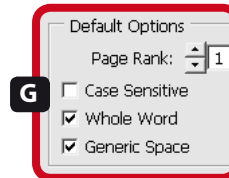
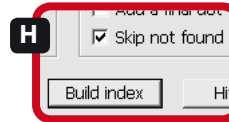
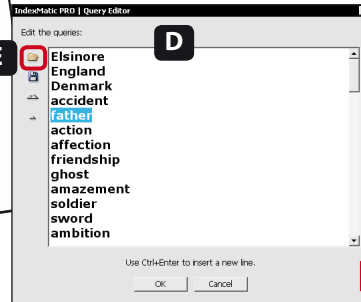
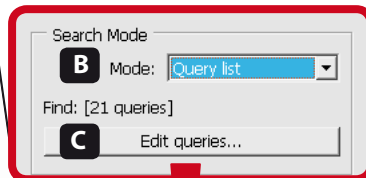
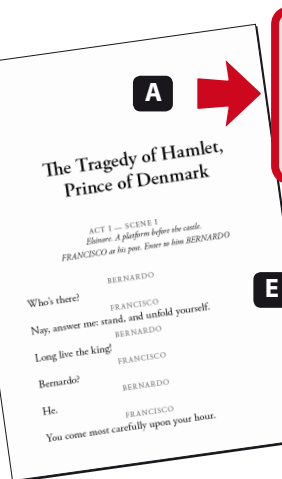
- In Automatic mode the **matching options** are disabled (and not used) but the **page rank** is available.
- Expressions that contain **spaces** cannot be found through the Automatic mode. In order to index all strings having a specific character style, you should rather send a **regex query**  $\rightarrow$  **19**, such as: `/.+/`

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025




Search Options



1	accident	79, 97, 150
2	action	31, 65, 80, 82, 87, 94, 110, 1
3	affair	11, 17, 102, 129
4	affection	24, 27, 147
5	affliction	80, 103, 143
6	amazement	103, 117
7	ambition	62-63, 88, 110, 133
8	beast	16, 36, 71, 132, 151, 175
9	beauty	24, 65, 83
10	blood	7, 23, 28, 34-37, 45, 71, 90,
11	body	16, 24, 33, 36-37, 88, 94, 115
12	breath	14, 83, 105, 121, 177, 184, 18
13	castle	1, 10, 49, 78, 87, 107, 122, 1
14	complexion	30, 71, 176
15	confession	64, 78, 151
16	conscience	77, 80, 82, 141, 147, 174, 187
17	death	8, 10-11, 14, 31, 35, 50, 52,
18	Denmark	1, 5, 12-13, 15, 24, 33, 35, 3
19	desire	15, 24, 31, 40, 141, 172
20	devil	77, 80, 93, 116, 120, 141, 165
21	earth	8-9, 14, 16, 22, 37, 42, 61, 6
22	Elsinore	1, 17, 63, 68, 75, 133
23	England	86, 107, 121, 129-130, 146, 14
24	Faith	17, 40-41, 45, 61, 67, 91, 96,
25	Father	7, 11-19, 22, 25, 31, 34-35, 4
26	fault	14, 30, 109
27	fellow	41, 87, 94, 159, 161, 164-165
28	fortune	30, 61-62, 73-74, 81, 90, 97,
29	friendship	63
30	Ghost	4-5, 8-9, 30-31, 33-36, 38-42,
31	grace	8, 15, 25, 30, 42, 52, 112, 11
32	grave	39, 60, 122, 136, 142, 156, 16
33	grief	10, 14, 49, 86, 97, 138, 142,
34	hate	49
35	heart	2, 12, 14-16, 24, 37, 39, 56,
36	heaven	4-5, 8, 14, 16, 18, 25, 28, 31
37	infinite	30, 62, 65, 165
38	judgment	25, 49, 90, 115, 126, 138, 142
39	kingdom	10, 130, 144, 149, 192
40	love	15, 19, 27, 35-36, 43, 48-49,

## 13 TURNING A WORD LIST INTO AN INDEX

Suppose you've just finished laying out "Hamlet" **A** and the publisher wants you to carry out a subject index based on an handful of predefined topics—such as: *love, death, madness, Denmark...*

1. Run IndexMatic<sup>2</sup> and choose 'Query list' **B** in the **Search Mode** panel.
2. The **Query Editor** **D** automatically opens (otherwise, click **Edit queries...** **C**). Here you can directly edit a list of topics, or import a word list from an existing plain text file **E** by clicking the  button →**14**.
3. Press **OK** to close the Query Editor. Note that the Search Mode panel indicates the **number of queries** to be processed **F**. Adjust the **matching options**, the **page rank** **G** and other desired settings. When everything is ready, click **Build index** **H**.

- When supplying a simple word list, it is usually a good option to enable **Whole Word** so you **prevent partial matches** like: **bullet, bulletin, bullion...**
- Items in a word list are not subject to the **Alphabet's** rules →**9**. Each item can contain foreign characters, inner spaces, etc. Hence a 'word list'—and more generally a 'query list'—might actually be used to achieve any kind of indexes.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Search Options

The screenshot shows the 'IndexMatic PRO | Query Editor' window. The main area contains a list of queries, each starting with two slashes (//) indicating a comment. The queries are: // Proper names, Gödel, Cantor, Turing, Hilbert, // Plural forms, //, /samples?/ => sample, /child(ren)?/ => child, //, // URL pattern, and a complex regex pattern. Annotations A, B, C, and D point to specific parts of the interface: A points to the main query list, B points to the file icons, C points to the zoom controls, and D points to the OK/Cancel buttons. Red arrows on the right side of the window point from the queries to explanatory text.

Open/save the list. **B**

Increase/decrease the type size. **C**

**A**

**D**

These comments won't be parsed as actual queries.

Basic 'token-based' queries. →16

Simple 'regex-based' queries. →19

A more complex pattern. →19

Use Ctrl+Enter to insert a new line.

OK Cancel

## 14 USING THE QUERY EDITOR

The **Query Editor** is a windowed interface available from the 'Query list' mode and the 'Edit queries...' button. It allows you to specify, edit, and manage a list of queries to be processed by the query engine →15.

1. You can directly type, copy, cut, and paste items in the main area **A**. (On some platforms you need to press **Ctrl+Enter** to insert new lines.)
2. Click the button to load a text file. Click the button to save the current list as a text file. **B**
3. The and buttons respectively increase and decrease the type size in the edit area. **C**
4. Click **OK** to record the changes you made in the list and close the editor. Click **Cancel** to ignore the changes (the list is then restored to its previous state before the window is closed). **D**

- Since IndexMatic v. 2.025, any line beginning with two slash signs (//) is considered **comment**. Comments have no effect on query processing, you can use them to make the query list more readable.
- Special comments in the form: `// topic1 => topic2` are used to emulate cross-references →20
- **Empty lines** are automatically removed from the list.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Advanced Queries

Q U E R Y

what you  
SEARCH



what you  
INDEX

**A**

**B**

castle

=>

castle

Not needed (implicit indexing term.)

Gödel  
manor  
fort

=>

Gödel, K.  
castle  
castle

Explicit rule (topic-rewriting, pooling.)

# 15

## ABOUT THE QUERY SYNTAX

1. What makes IndexMatic<sup>2</sup> highly flexible is its **query interpreter**. Basically, each query to be processed is split in two parts: the **SEARCH** side **A** and the **INDEX** side **B**. Most of the time this dichotomy remains implicit and invisible to the user. For example, when a query is supplied as a simple word—say **castle**—the system assumes that **castle** is both the searched expression *and* the heading that needs to be reported in the final index. Behind the scene, this simple query is actually seen as: **castle=>castle**.
2. But in some situations this default mechanism is unsuitable. First you may want to **reformat** the search key: **Gödel=>Gödel, K.** Furthermore, you may have to capture **variant forms**, or to **pool together** different words (*castle, manor, fort*) in a **single topic** (*castle*). Then you will use the => operator, as illustrated below...

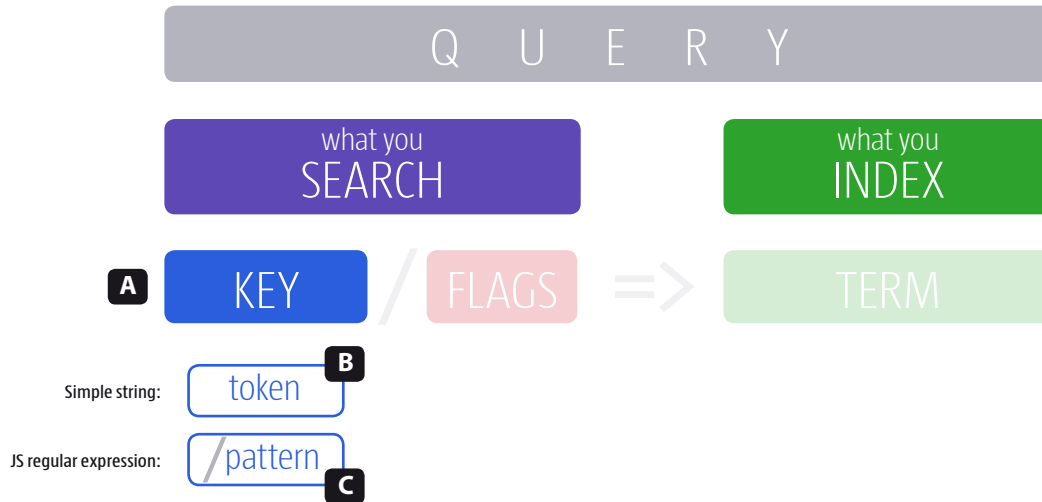
- The full query syntax is available in both **Query list** and **Single query** modes.
- The **rewriting operator** must be exactly typed as shown: => (An equal sign followed by a greater-than character.) Note that you are allowed to insert extra spaces **around** any operator.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Advanced Queries



# 16

## QUERY SYNTAX: THE 'KEY'

1. The **KEY** **A** is the **only required** element in a query. In most cases this simply is the word or expression you are searching for. We call it a **token** **B**. If you don't use extra operators, any string will act as a simple token and forms a complete query.
2. You can also supply a **regular expression**. To do so, insert a slash (/) at the beginning of the **pattern** **C**. E.g.: `/dog|cat|snake` will find any of these three words, while `/cats?` grabs both `cat` and `cats`. An (optional) ending slash is allowed: `/cats?/`. (For further details about regex-driven queries →**19**)
3. Depending on your **Default settings** →**11**, a KEY—token or pattern—may be **case sensitive** or not. It may also support **generic space** and similar options →**17**.

- Since the symbol / has a special meaning on the KEY side, you might have to use the escape sequence `\/` to introduce this character as itself.
- Inversely, any **pattern** requires a **starting-slash**, otherwise it is parsed as a **token**—disregarding any regex operator or metacharacter.

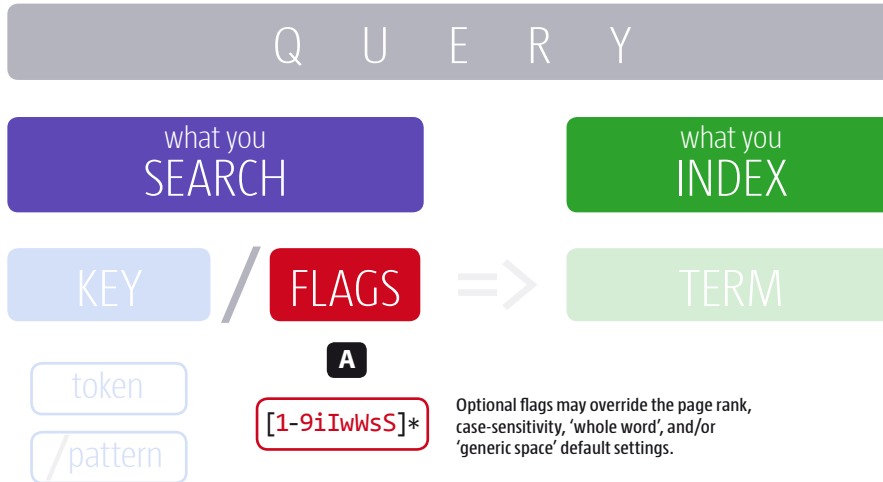


# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Advanced Queries



## 17

### QUERY SYNTAX: THE 'FLAGS'

1. Optional **FLAGS** <sup>A</sup> allow you to specify how a singular KEY is interpreted by the query engine. To introduce one or several FLAG characters, use a slash (/) at the beginning of the sequence. You can combine FLAGS in any order: **cat/3w,/Cat|Dog/I2**, etc. The allowed characters are: **1 2 3 4 5 6 7 8 9 i I w W s S**
2. Any digit from **1** to **9** is parsed as a **page rank** flag. Use it to override the default page rank →**10**.
3. **i** makes the query **case-insensitive**, **I** makes the query **case-sensitive**, disregarding the defaults →**11**.
4. **w** forces a **whole word** query, **W** forces a **non-whole word** query, disregarding the defaults →**11**.
5. **s** enables the **generic space** feature, **S** disables the **generic space** feature, disregarding the defaults →**11**.

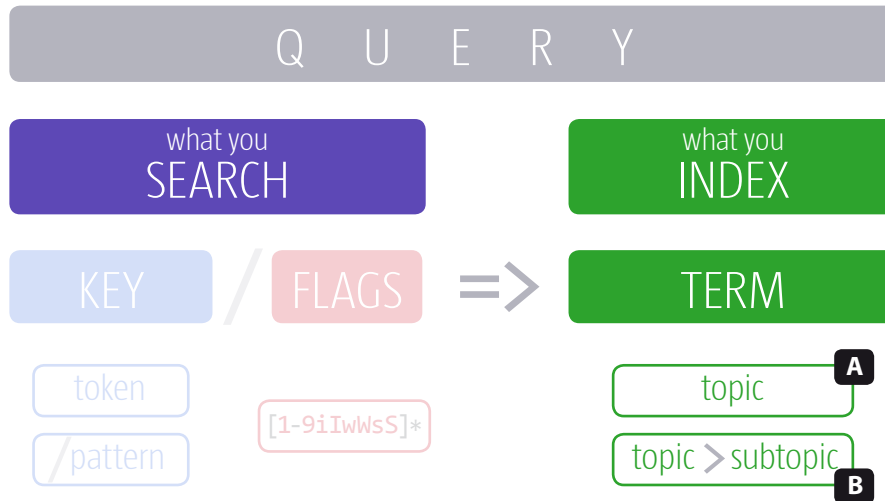
- The main function of a flag is to **override** the default settings when necessary. For example, assuming your global page rank is 3, you can apply a lower condition in a specific query: **beauteous/1**
- Note that the **s** and **S** flags are meaningless if the key does not contain any space.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Advanced Queries



Supplying an explicit TERM opens up many possibilities: grouping multiple keys under a single topic, rewriting a complex pattern, managing the topic>subtopics relationship.

## 18 QUERY SYNTAX: THE 'TERM'

The **TERM**—rightmost part of the query syntax—reflects the **actual** string to be rendered in the index. Depending on the KEY, the query engine may implicitly determine one—or several—default TERM(s). However, any query can **explicitly decouple** the TERM from the KEY using the **rewrite** operator (**=>**).

1. In its basic form, an explicit TERM is a string, such as **animal**. It then “takes the place” of the KEY and serves as a **topic** in the final index <sup>A</sup>. For example: **dog=>animal** tells the interpreter to connect the token **dog** to the topic **animal**. (Of course you can re-use a same TERM in different queries.)
2. The TERM field also supports a special operator, **>**, that lets you both create a **subtopic** and specify a **parent topic** <sup>B</sup>. For example: **dog=>animal>dog** will index **dog** as a subtopic *under* the topic **animal**.

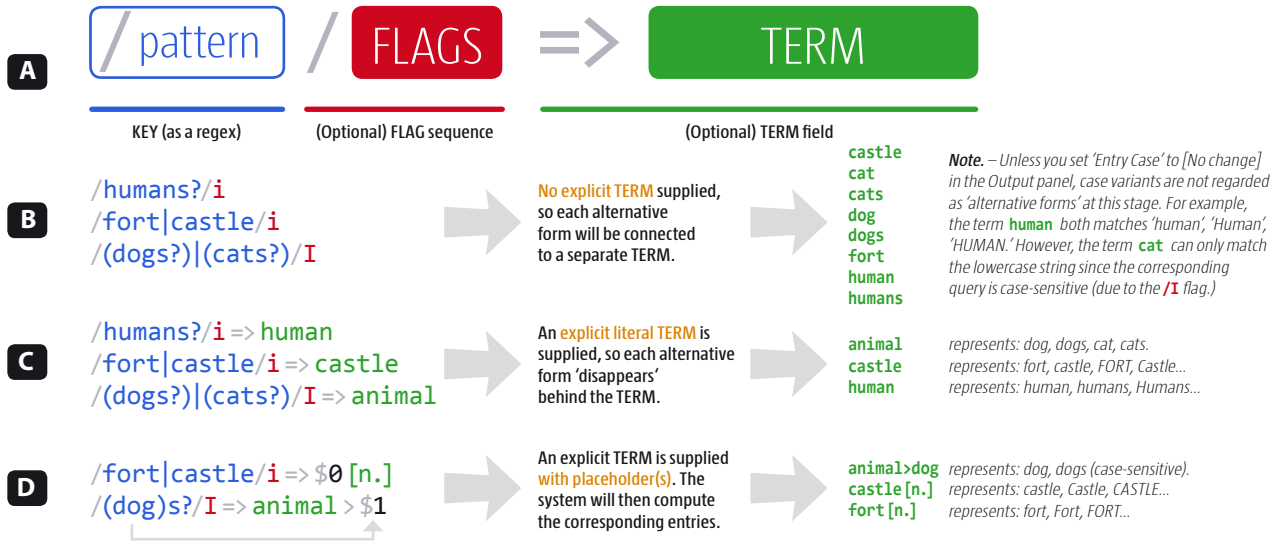
- Imagine a topic as a top-level heading that is allowed to **parent** other child strings (subtopics). IndexMatic<sup>2</sup> only allows **two hierarchical levels**.
- Since the symbol **>** has a special meaning on the TERM side, you may have to use the escape sequence **\>** to introduce this character as itself.
- How topic levels are formatted is discussed in the ‘Output Preferences’ section →23

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Advanced Queries



# 19

## REFINING REGEX-BASED QUERIES

1. Inherently a regular expression will match **different forms** in the text. Hence, when processing a regex-based query **A**, IndexMatic<sup>2</sup> assumes that **each found form** should belong to a separate TERM **B**—provided that no explicit TERM is supplied.
2. Now, if the query contains an **explicit TERM** **C**, the rewriting operator ‘aggregates’ the matches in the **corresponding topics**.
3. Finally, the query engine allows you to insert **placeholders** (**\$0**, **\$1**, ..., **\$9**) in the TERM field. These represent any **matched substring** that results from a **capturing parenthesis**. **\$0** serves as the global match, while **\$n** serves as the *n*<sup>th</sup> capture in the regular expression (counting left parentheses) **D**.

- Thanks to the ‘Entry Case’ option of the Output panel, the indexing TERMS can be **post-formatted**. For further details →**24**
- Older versions of the script were supporting **\$** as a shorthand of **\$0**. Now this symbol, when used alone, represents the KEY itself (i.e. the token or pattern). For example, `/dogs?/ => $` is parsed as: `/dogs?/ => dogs?`

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Advanced Queries



**B** `// castle`  $\Rightarrow$  See also fort.  
`// animal > dog`  $\Rightarrow$  See canine.

# 20

## EMULATING CROSS-REFERENCES

1. A special comment syntax is allowed to indicate a **cross-reference** **A**: `// TERM  $\Rightarrow$  REFERENCE`  
This pattern is useful to inject a FAKE TERM in the index, and to visually link it to another location through the REFERENCE text. (Note that no KEY is used in this model.)
2. Any valid TERM syntax is allowed **B**, including the **topic>subtopic** form  $\rightarrow$ 18, but of course you cannot use a **\$n** placeholder in such context.
3. Any string can serve as a REFERENCE text, usually in the form "See. . ." or "See also. . ."  
The interpreter does not control the possible existence of the term(s) mentioned here.

- A cross-reference is not really sent to the query system. The 'fake term' is just treated as if it was an actual resulting term, and the reference text is then displayed  $\rightarrow$ 23 instead of a page location.
- Both the comment marker (`//`) and the  $\Rightarrow$  operator are necessary to create a valid cross-reference. If the  $\Rightarrow$  operator is omitted, the 'query' is parsed as a regular comment  $\rightarrow$ 14.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Advanced Queries

Operator	Use
<code>\</code>	Used to escape other special characters: <code>. \ / + * ? [ ^ ] \$ ( ) { } = ! &lt; &gt; : -  </code>
<code>*</code>	Matches the preceding element 0 or more times.
<code>+</code>	Matches the preceding element 1 or more times.
<code>?</code>	Matches the preceding element 0 or 1 time.
<code>{n}</code>	Matches exactly <i>n</i> occurrences of the preceding element.
<code>{n,}</code>	Matches at least <i>n</i> occurrences of the preceding element.
<code>{n,m}</code>	Matches at least <i>n</i> and at most <i>m</i> occurrences of the preceding element.
<code>x y</code>	Alternative. Matches either <i>x</i> or <i>y</i> .
<code>(x)</code>	Capturing parenthesis. (See <a href="#">\$</a> →19.)
<code>(?:x)</code>	Non-capturing parenthesis.
<code>x(?:y)</code>	Positive lookahead.
<code>x(?:!y)</code>	Negative lookahead.
<code>[xyz]</code>	Character set. Matches any of the enclosed characters.
<code>[^xyz]</code>	Complemented character set. (The symbol <code>^</code> is not supported in any other context.)

(1) Since the following characters are removed from the text stream before searching, they are *never matched* by any metacharacter: paragraph return, column break, frame break, page breaks, page number, text variables, End Nested Style Here, Non Roman Special Glyph, Discretionary Hyphen, Discretionary Line Break, Non-joiner, Zero Width Joiner, and all anchors and markers.

Warning: `\w`, `\W`, `\l`, `\L`, `\m` and `\M` are all *alphabet-relative*. This is a non-standard implementation. Also, note that the 'uppercase' metacharacters are `\m` and `\M` (instead of the usual `\u` and `\U`).

Positive and complementary character sets both support the following metacharacters: `\w`, `\W`, `\l`, `\L`, `\m`, `\M`, `\uHHHH`, and `SPACE` in 'Generic Space' queries.

Metachar.	Use
<code>SPACE</code>	If 'Generic Space' is enabled, matches any space character. (See table →11.) Escape sequence: <code>\SPACE</code>
<code>.</code> (dot)	Matches any single character <sup>1)</sup> .
<code>\d</code>	Matches a digit character. Equivalent to <code>[0-9]</code> .
<code>\D</code>	Matches any non-digit character, i.e. <code>[^0-9]</code> .
<code>\w</code>	Matches any character of the current Alphabet. (See →9.)
<code>\W</code>	Matches any character that doesn't belong to the current Alphabet. (See →9.)
<code>\l</code>	Matches any lowercase letter of the current Alphabet. (See →9.)
<code>\L</code>	Matches any character that is not a lowercase letter of the current Alphabet. (See →9.)
<code>\m</code>	Matches any uppercase character of the current Alphabet. (See →9.)
<code>\M</code>	Matches any character that is not an uppercase letter of the current Alphabet. (See →9.)
<code>\t</code>	Matches the TAB character.
<code>\s</code>	Equivalent to: <code>[ \t\u00A0\u2028\u2029]</code> . (Instead, consider to use generic <code>SPACE</code> .)
<code>\uHHHH</code>	Matches the Unicode character U+HHHH
<code>\b</code> <code>\B</code> <code>\s</code> <code>\u</code> <code>\f</code> <code>\n</code> <code>\r</code>	Not used in patterns.

A

# 21

## SPECIAL CHARACTERS IN REGULAR EXPRESSIONS

1. IndexMatic<sup>2</sup> supports **ECMAScript Regular Expressions**, whose syntax and semantics are fully described in the **ECMA-262** Language Specification (section 15.10): <http://www.ecma-international.org/publications/standards/Ecma-262.htm> For a good overview of JavaScript regex and practical examples, we also invite you to visit the **MDN's Guide to Regular Expressions**: [http://developer.mozilla.org/en/JavaScript/Guide/Regular\\_Expressions](http://developer.mozilla.org/en/JavaScript/Guide/Regular_Expressions)
2. In addition, IndexMatic<sup>2</sup> provides some special metacharacters **A** that **supersedes** the original specification in order to make the syntax easier and to provide more relevant results. (See also next page.)

- The `^` and `$` symbols are not used in patterns. In a **TERM** field, `$` represents the original key (`$`) or a captured match (`$0`, `$1`, `$2`, ..., `$9`). Escape sequence: `\$`.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## Advanced Queries

Warning: when using a metacharacter that *discriminates case*, such as `\p{Lu}`, make sure your query is *case sensitive* → **11**

While the negative metacharacter `\P{...}` is *not implemented*, you can use a complementary set in the form: `[^\p{xx}]`.

### A Unicode Properties

Property	Use	Property	Use
<code>\p{Ll}</code>	Lowercase letter. E.g.: é, α, ā, œ, ç...	<code>\p{Ps}</code>	Open punctuation. E.g.: (, [, {...
<code>\p{Lu}</code>	Uppercase letter. E.g.: A, É, Ð, Ω, NJ...	<code>\p{Pe}</code>	Close punctuation. E.g.: ), ], }...
<code>\p{Lt}</code>	Titlecase letter. E.g.: Dž, Lj, Nj...	<code>\p{Pi}</code>	Initial punctuation. E.g.: «, ’, ...
<code>\p{Lm}</code>	Modifier letter. E.g.: ¨, ¨, ¨...	<code>\p{Pf}</code>	Final punctuation. E.g.: », ’, ...
<code>\p{Lo}</code>	Letter without case. E.g.: 2, ll, ð...	<code>\p{Pc}</code>	Connector punctuation. E.g.: _ , ~, } ...
<code>\p{M}</code>	Any mark.	<code>\p{Po}</code>	Other punctuation. E.g.: !, ;, #...
<code>\p{Mn}</code>	Non-spacing mark.	<code>\p{S}</code>	Any symbol.
<code>\p{Mc}</code>	Spacing combining mark.	<code>\p{Sm}</code>	Math symbol. E.g.: +, <, ↔...
<code>\p{Me}</code>	Enclosing mark.	<code>\p{Sc}</code>	Currency symbol. E.g.: \$, €, £, ¤...
<code>\p{N}</code>	Any number.	<code>\p{Sk}</code>	Modifier symbol. E.g.: ^, ¯, ¨, ...
<code>\p{Nd}</code>	Decimal digit. E.g.: 1, 2, ¾...	<code>\p{So}</code>	Other symbol. E.g.: †, §, ©...
<code>\p{Nl}</code>	Letter number. E.g.: viii, ©, ¢...	<code>\p{Z}</code>	Any separator.
<code>\p{No}</code>	Other number. E.g.: ², ½, ¼...	<code>\p{Zs}</code>	Space separator.
<code>\p{P}</code>	Any punctuation.	<code>\p{C}</code>	Any other character.
<code>\p{Pd}</code>	Dash punctuation. E.g.: -, —, ~...		

### B

#### Additional GREP shortcuts

Op.	Character
<code>~8</code>	Bullet •
<code>~e</code>	Ellipsis ...
<code>~7</code>	Paragraph Symbol ¶
<code>~6</code>	Section Symbol §
<code>~2</code>	Copyright ©
<code>~r</code>	Registered Trademark ®
<code>~d</code>	Trademark Symb. ™
<code>~_</code>	Em Dash —
<code>~=</code>	En Dash –
<code>~{</code>	Double Left Quot. “
<code>~}</code>	Double Right Quot. ”
<code>~[</code>	Single Left Quot. ‘
<code>~]</code>	Single Right Quot. ’
<code>~"</code>	Straight Dble Quot. "
<code>~'</code>	Straight Single Quot. '
<code>~S</code>	Nonbreaking Space
<code>~5</code>	Fixed Width Nonbrk. Sp.
<code>~i</code>	Indent To Here
<code>~y</code>	Right Indent Tab
<code>~f</code>	Flush Space
<code>~&gt;</code>	En Space
<code>~m</code>	Em Space
<code>~3</code>	Third Space
<code>~4</code>	Quarter Space
<code>~%</code>	Sixth Space
<code>~/</code>	Figure Space
<code>~.</code>	Punctuation Space
<code>~&lt;</code>	Thin Space
<code>~ </code>	Hair Space
<code>~~</code>	Tilde (escape seq.)

# 22

## ADDITIONAL METACHARACTERS

- Since v. 2.025, IndexMatic<sup>2</sup> supports the `\p` metacharacter in the form `\p{xx}`, where `xx` refers to a **Unicode property**. The available properties are listed in table **A**. This feature allows to search characters by property, **disregarding the current Alphabet** → **9**. For example, `\p{Ll}` matches any Unicode lowercase letter whereas `\l` only matches Alphabet's lowercase letters. Hence: `[a-z] ⊆ \l ⊆ \p{Ll}`
- The query engine also supports a special **subset of GREP metacharacters** **B**.
- The metacharacters above can be used alone, or within a character set. E.g.: `[_\p{Ll}]`, `[~3~4~%]`, `[^\p{P}]`, etc.

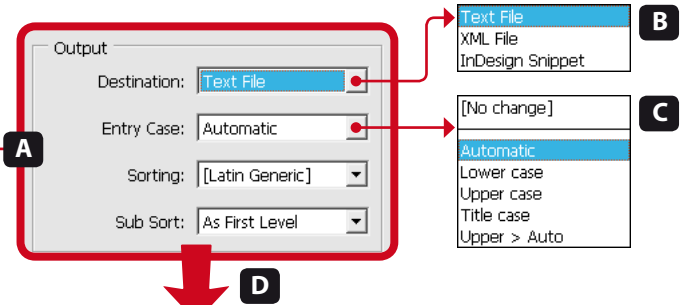
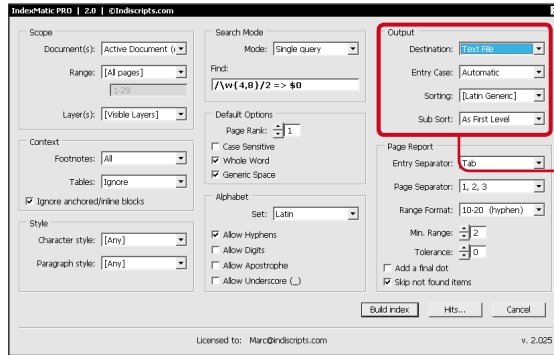
■ **Keep in mind that the regular expression syntax used in IndexMatic<sup>2</sup> is not identical to InDesign GREP syntax—although there are considerable similarities between the two languages.**

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Output



Each line in the text file represents a distinct 'index entry'. A tab (or other Entry Separator →26) separates each topic from the corresponding set of page numbers. Subtopics (*Denmark, Elsinore, England*) are always grouped under their parent topic (*PLACES*). An indentation is applied to every group of subtopics.

```
46 passion 48, 74, 76, 87, 90, 97, 117, 143, 175
47 peace 4, 22, 65, 114, 132, 149-150, 173, 175,
48 PLACES
49 » Denmark 1, 5, 12-13, 15, 24, 33, 35, 37-39,
50 » Elsinore 1, 17, 63, 68, 75, 133
51 » England 86, 107, 121, 129-130, 146, 163, 172
52 power 10-11, 36, 50, 73, 77, 83, 130
53 reason 12, 14, 16, 30, 32, 57, 60, 63, 65, 85,
```

## 23 RENDERING THE INDEX IN A PLAIN TEXT FILE

The **Output** panel **A** is the place to specify how the final index will be rendered.

1. In the **Destination** list, select 'Text File' (default **B**) to get the resulting index in an editable **plain text file**. The other options are 'XML File' →27 and 'InDesign Snippet' (CS4/CSS only) →28.
2. In the **Entry Case** list, select 'Automatic' (default **C**) to let IndexMatic<sup>2</sup> automatically determine the case of final entries (for further details about the underlying algorithm →24). Except for particular format, this option is generally the most relevant.
3. In the **Sorting** option list, select '[Latin Generic]' (default) to obtain properly sorted entries **D**. Finally, press **Build index**.

- Each index file is created with a new unique **timestamp name**, e.g. *indexmatic\_2011-07-01\_20h52\_21*, and located in the indexed-document folder. (If the document has not been saved to disk yet, the index file is placed on the saved desktop.)
- The **TRY version** only allows 'Text File' destination.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Output

dog	Dog	DOG
1	2	3

The 'Entry Case' field allows you to 'post-format' the case of the final index entries. The table shows the results for different queries addressed to the same document.

- A** implicitly creates the TERM **dog** (lowercase) so the system does not track case variants and always renders a single item.
- B** computes the TERM(s) via **\$0**, so the items are rendered separately in '[No Change]' mode. 'Automatic' promotes the most used case.
- C** is case-sensitive: the query can only find the token **Dog** (on page 2). Then the implicit TERM (**Dog**) is formatted.
- D** illustrates the use of an explicitly rewritten TERM (**Animal**).
- E** is similar to **B** except the computed TERM contains a parent topic (**Animal**). Note how the option 'Upper > Auto' renders the items.

	SAMPLE QUERY	<b>dog/i</b> <b>A</b>	<b>/do\w/i=&gt;\$0</b> <b>B</b>	<b>/Dog/I</b> <b>C</b>	<b>dog/i=&gt;Animal</b> <b>D</b>	<b>/dog/i=&gt;animal&gt;\$0</b> <b>E</b>
	ENTRY CASE					
[No Change]	dog 1-3	dog 1-3	Dog 2	Animal 1-3	animal dog 1 Dog 2 DOG 3	
Automatic	dog 1-3	dog 1-3	Dog 2	Animal 1-3	Animal dog 1-3	
Lower Case	dog 1-3	dog 1-3	dog 2	animal 1-3	animal dog 1-3	
Upper Case	DOG 1-3	DOG 1-3	DOG 2	ANIMAL 1-3	ANIMAL DOG 1-3	
Title Case	Dog 1-3	Dog 1-3	Dog 2	Animal 1-3	Animal Dog 1-3	
Upper > Auto	dog 1-3	dog 1-3	Dog 2	Animal 1-3	ANIMAL dog 1-3	

## 24

### UNDERSTANDING THE 'ENTRY CASE' OPTIONS

- Each query TERM →**18** is temporarily stored in a structure that registers case variants. So the script is able to differentiate explicit TERMS like **Animal** and **animal**. Computed TERMS follows the same rule. For example, while the query **/dogs?/i** may match **dog, Dogs, DOG, dogs**, etc., IndexMatic<sup>2</sup> internally creates two templates, {dog} and {dogs}, and maintain a counter for case variants.
- Then TERMS are rendered according to the **Entry Case** option: '[No Change]' means that each case variant is preserved as a **distinct item**. 'Automatic' collects variants and applies the **most used case** to the final entry. 'Lower/Upper/Title Case' applies the corresponding format (whatever the tracked variants). 'Upper > Auto' applies 'Upper Case' to the topic and 'Automatic' to the subtopic **if present**, otherwise it acts like 'Automatic.'

- When a **pattern** does not contain any special regex operator, IndexMatic<sup>2</sup> mutely optimizes the query and converts this pattern into a simple **token**. For example, **/dog/i** is considered **dog/i** (without starting-backslash). Hence the system creates an implicit TERM (**dog**) which does not track case variants (see **A**). To force case-variant tracking, use a placeholder: **/dog/i => \$0**



# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Output

Output

Destination: Text File

Entry Case: Automatic

Sorting: [Latin Generic]

Sub Sort: As First Level

[No Sort]

As First Level

Numeric Sequence

**C**

Remember that IndexMatic<sup>2</sup> basically supports Latin-based languages → 9. If your topics involve characters from another alphabet—such as Cyrillic—the script cannot properly sort them.

[No Sort]

[Latin Generic]

Albanian

Basque

Bosnian

Breton

Croatian

Czech

Danish

Dutch (IJ=Y)

Dutch (Modern)

English

Esperanto

Estonian

Faroese

Filipino (Modern)

Finnish (Multilingual)

Finnish (Official)

French

Frisian

German (Decompose umlaut)

German (Dictionary)

Greenlandic

Hungarian

Icelandic

Indonesian

Irish Gaelic

Italian

Kurdish (Hawar)

Latvian

Lithuanian

Luxembourgish

Malay

Maltese

Moldavian

Norwegian

Polish

Portuguese

Portuguese

Romanian

Serbian

Slovak

Slovene

Spanish (Modern)

Spanish (Traditional)

Swedish (Modern)

Swedish (V=W)

Turkish

Turkmen

Welsh

**EXAMPLE: Ordering Norwegian topics.**

Sorting:  
[Latin Generic]

**A**

åbner  
ære  
altså  
brønder  
brød  
fjord  
bryst  
fjord  
jeg  
ordet  
vild

Sorting:  
Norwegian

**B**

altså  
bryst  
brønder  
brød  
fjord  
jeg  
ordet  
vild  
ære  
øje  
åbner

# 25

## ORDERING THE FINAL INDEX ENTRIES

At the bottom of the Output panel are the **Sorting** and **Sub Sort** dropdown lists. The first offers a large number of language items. Each corresponds to a **'collation algorithm'** which sets specific alphabetical-ordering rules.

1. Select **'[No Sort]'** to completely discard these features. Final entries are then left unsorted, which might be useful if you need to preserve the original order of explicit query TERMS supplied through the Query Editor.
2. Select **'[Latin Generic]'** **A** to get your topics the most appropriately sorted whatever the underlying language(s). This option is especially relevant when addressing foreign words or multilingual documents.
3. Select a **more specific algorithm** **B** to tell IndexMatic<sup>2</sup> to apply the rules of the target-language.
4. **Subtopics** are ordered according to the **Sub Sort** option **C**: **'[No Sort]'**, **'As first level'** or **'Numeric Sequence'**.

- When **[No Sort]** is selected during an 'Automatic' search process or applied to computed TERMS, you cannot make any assumption about the ordering of the final data.
- If the index owns **subtopics** → 18, they are sorted by default as the first-level topics: **Sub Sort > 'As First Level'**. Use the **'Numeric Sequence'** option to apply a numerical sort instead.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Output

The screenshot shows the IndexMatic PRD 2.0 interface. The 'Page Report' panel is highlighted with a red box and labeled 'A'. It contains the following settings:

- Entry Separator: Tab
- Page Separator: 1, 2, 3
- Range Format: 10-20 (hyphen)
- Min. Range: 2 (labeled 'E')
- Tolerance: 0 (labeled 'F')
- Add a final dot (labeled 'G')
- Skip not found items (labeled 'H')

Arrows point from these settings to their respective output examples:

- 'Entry Separator' points to 'B' (Tab, Comma + Tab, Colon + Tab, Space, Comma + Space, Colon + Space).
- 'Page Separator' points to 'C' (1 2 3, 1, 2, 3, 1; 2; 3, 1; 2; 3, 1 | 2 | 3, 1 • 2 • 3).
- 'Range Format' points to 'D' ([No Range], 10-20 (hyphen), 10-20 (nonbreaking hyphen), 10-20 (en dash), 10—20 (em dash), 10\_20 (underscore), 10/20 (slash)).
- 'Min. Range' points to 'E' (2).
- 'Tolerance' points to 'F' (0).
- 'Add a final dot' points to 'G' (checkbox).
- 'Skip not found items' points to 'H' (checkbox).

## 26 FORMATTING THE PAGE NUMBERS

The **Page Report** panel **A** specifically regards how **entries** and **page numbers** are rendered in the index:

- Entry Separator B**: string inserted between entries and page locators (default is Tab).  
**Page Separator C**: string inserted between page numbers or page ranges (the default model is: **1, 2, 3**).
- Range Format D**: specifies how page ranges are formatted. Select "[No Range]" to prevent pages from being ranged. **Min. Range E** (2-10): minimum number of **consecutive page numbers** required in a range. **Tolerance F** (0-5): number of allowed **missing page numbers**—"holes"—in a range. Check the **Add final dot G** to have each line ending with a dot point.
- Skip not found items H**: check this box to prevent **not found** entries from being displayed. Otherwise, an **EM DASH (—)** serves as empty marker: **NotFoundTopic —**

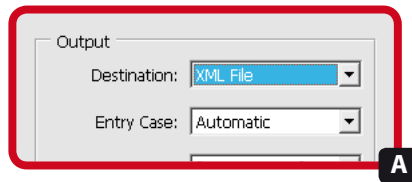
■ **Page names** are numerically sorted when possible. If the document(s), or some sections, use special **numbering styles**—such as "i, ii, iii, iv..." or "a, b, c..."—then the corresponding pages are reported under their own names **at the beginning** of the locator. E.g.: **topic vi, vii, xi, 53, 130-131...** (These special page names are not ranged.)

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Output



A

**topic element**  
(may contain a 'hits' attribute, if available)

**locations element**  
(the Page Rank condition is stored as an attribute)

**subtopic elements**  
(same structure as topic)

```
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>
36  </topic>
37  <term>madness</term>
38  </term>
39  </index>
```

B

**Header**

**term element**  
(may contain a 'variants' attribute if case variants are tracked)

The **locations** element may be empty. Here the query has not been satisfied (due to the high page rank condition), but 'not found entries' need to be rendered (due to the output preferences).

## 27 XML EXPORT

1. In the **Output** panel, select the destination: 'XML File' **A**.
2. Adjust as needed the other preferences: **Entry Case** → **24**, **Sorting** → **25** and **Page Report** → **26**. These settings have the same meaning and act the same way in XML export.
3. Click **Build index** to generate the file **B**. The resulting XML structure is **more or less verbose** depending upon the global page rank, the 'Entry Case' option, or other settings that control the query system. Here is the minimal structure of an element:

```
<topic><term>...</term><locations>...</locations></topic>
```

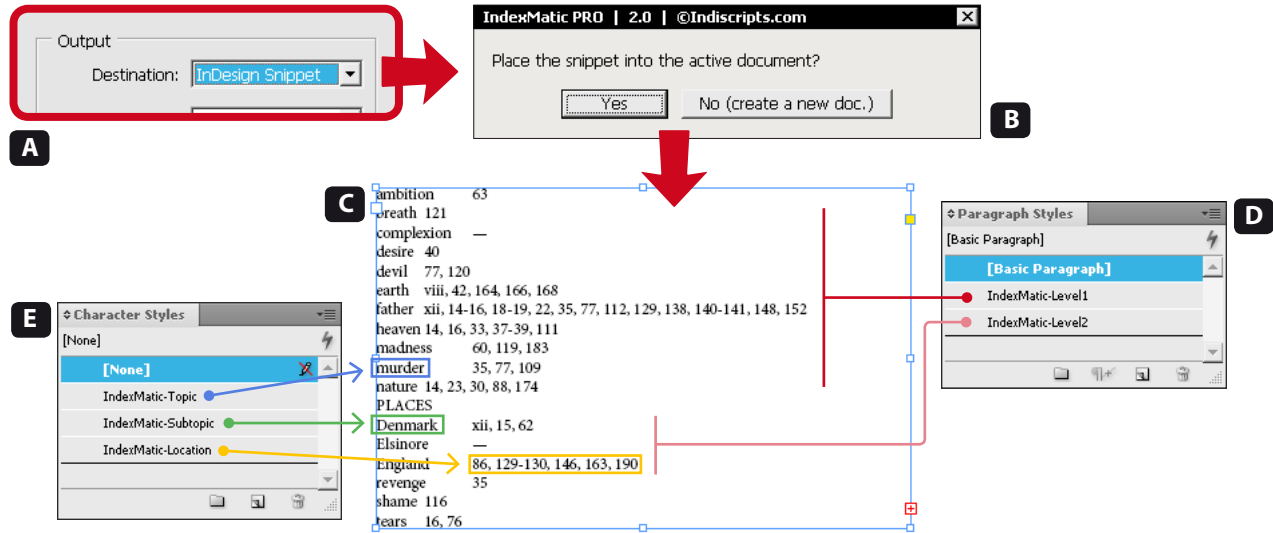
- XML Export is disabled in the **TRY** version.
- Each XML file is created with a new unique **timestamp name**, e.g. `indexmatic_2011-07-01_20h52_21.xml`, and located in the indexed-document folder. (If the document has not been saved to disk yet, the index file is placed on the desktop.)

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Output



## 28 PLACING THE INDEX AS A 'SNIPPET' (CS4/CS5)

1. In the **Output** panel, select the destination: 'InDesign Snippet' **A**.
2. Adjust as needed the other preferences: **Entry Case** → **24**, **Sorting** → **25** and **Page Report** → **26**. These settings have the same meaning and act the same way with snippets.
3. Click **Build index**. You have the option to get the result placed into the active document or into a new one **B**. The snippet is then loaded through the **place cursor**. Just click in your document to 'wake up' the underlying text frame.
4. Note that no particular enrichment is applied to the entries **C**. Instead, IndexMatic<sup>2</sup> creates a set of **Paragraph** & **Character styles** **D E** that allow you to format each element at your own convenience.

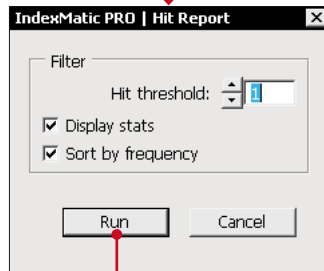
- This feature is disabled in the **TRY version**, and not implemented in **InDesign CS3**.
- An interesting point is that once you have refined **IndexMatic's default styles**, they are **preserved** when you generate and place a new snippet in the same document.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Output



ITEMS	HITS
1	-----
2	-----
3 the	1142
4 and	964
5 you	550
6 HAMLET	465
7 that	404
8 not	313
9 lord	310
10 his	296
11 this	296
12 but	270
13 with	267
14 for	252
15 your	242
16 what	216
17 KING	202
18 in	196
19 have	180
20 will	169
21 HORATIO	158
22 are	131
23 all	122
24 CLAUDIUS	120
25 POLONIUS	119
26 our	118
27 QUEEN	118
28 shall	114
29 good	109
30 come	106
31 let	105
32 thou	105
33 they	104
34 LAERTES	104
35 now	98
36 more	96
37 GERTRUDE	96
38 from	95
39 there	95
40 her	91
41 How	90
42 thy	87
43 OPHELIA	87
44 was	85
45 most	82
46 like	80
47 would	79
48 well	78
49 know	77
50 ROSENCRANTZ	76
51 sir	75
52 them	74
53 tis	73
54 may	70
55 father	69
56 love	66
57 did	65
58 First	65
59 Enter	64
60 very	64
61 then	64
62 GUILDENSTERN	64
63 speak	63
64 which	63
65 here	62
66 hath	62
67 why	60
68 must	60
69 thee	59
70 give	58
71 where	58
72 should	58
73 man	57

Hamlet's word stats powered by IndexMatic<sup>2</sup> using the Automatic search mode →12.

## 29

### GETTING MATCHES & STATS BEFORE INDEXING

Since v. 2.025, IndexMatic<sup>2</sup> offers a **Hit Report** feature which you can use to collect matches without page numbers. This is useful to prepare a relevant word list, test queries or simply retrieve word stats.

1. Once your queries and settings are defined, press the **Hits...** button to open the **Hit Report** window **A**.
2. The **Hit threshold** field (1–100) allows to filter matches that don't occur at least *N* times in the entire Scope →4. The default value is 1, which inhibits the filter. Set a higher frequency to exclude rare terms from the report.
3. Check **Display stats** to include in the report the frequency of each resulting term. Uncheck this box to have terms displayed without stats (doing so does not disable the hit-threshold filter though).
4. Check **Sort by frequency** to order items by decreasing frequencies **B**. This bypasses the sorting options →25.

- Any Hit Report results in a plain text file created with a unique **timestamp** name, e.g. `indexmatic-hits_2011-10-04_20h52_21.txt`, and located in the indexed-document folder.

- Note that the **Hit threshold** filter acts **AFTER** query processing. Hence, if a match is ignored due to the **Page Rank** →10, it will not appear in the report whatever its frequency in the Scope.



## ABOUT PAGE CONTENT EXTRACTION

Depending on your settings—scope, context, styles, matching options...—IndexMatic uses different strategies to optimize the process of extracting page contents. In all cases, the script explores at some level the structure of the document(s) in order to properly identify which page ‘contains’ which piece of text. In InDesign, threaded frames, anchored objects, tables, and footnotes make it sometimes difficult to compute the correct page location of a ‘match.’ Here are some basic details on how IndexMatic deals with these issues.

### 1. THREADS AND BREAKS

When scanning the ‘scope’, IndexMatic first identifies the text frames located on the corresponding pages. A text frame is nothing but a visual container. It generally reflects partial contents of underlying text streams that flow independently: story, footnote(s), table cell(s). A single page may address multiple parts of multiple streams. Sometimes a text stream continues on another page, sometimes it spans different frames on the same page, sometimes it breaks and causes ‘overset.’ **The indexing process involves determining the actual boundaries of each contiguous text which is actually available on the page (or seen as such).** For example, when a hyphenated word spans two threaded frames, the processor should ignore this ‘visual break’ and assert that the full word simply occurs on the page—even if the end of the word actually stands on the next page. In contrast, if two frames are adjacent (on the same page) but correspond to distinct stories, the processor should not label respective text contents as being ‘threaded.’

### 2. LOOK BEHIND

Suppose you want to index the occurrences of “Alan Turing” in a document based on threaded text frames. If the string spans two pages—say “Alan ” on page A, “Turing” on page B,—IndexMatic must be able to assert that an occurrence of “Alan Turing” exists on page A.

To do so, the script ‘looks behind’ the current page and maintains a special buffer that contains the first characters of the next page. For performance reasons, **the size of this buffer is limited to 16 characters.** Hence, IndexMatic cannot always detect long matches across two pages.

### 3. FOOTNOTES & TABLES

Prior to v. 2.025, IndexMatic was considering **continued-footnote text** entirely located on the page owning the footnote marker. This issue has been solved.

Also, the script now reports the **correct location of cell contents** when the parent table spans two or more pages. However, IndexMatic still ignores any **nested tables**. Note that the table parser is distinct from the main algorithm and can lead to (significant) additional computing time.

### 4. PARAGRAPHS AND STYLE RANGES

Each time IndexMatic identifies a relevant text range on a given page—story range, footnote range, cell—it stores the underlying data in a special structure called ‘page run.’ **Each page run is then divided into ‘segments’ which correspond to the largest text units that a query can address.** No segment can be larger than the size of a paragraph. That’s why a ‘maximal’ query such as **/.\*+/** never match more than a paragraph content (excluding paragraph return).

Some segments are smaller than paragraphs. For example, if the user targets a specific character style, IndexMatic makes segments accordingly reflect the corresponding text ranges. In this case, a maximal query will not return results that extend beyond the style range.

However, when the user target a **character style group**, adjacent style ranges which belong to this group are merged into a unique segment, which allows to extract text content standing across multiple character styles within the group.

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



## LIMITATIONS, KNOWN ISSUES

### 1. LANGUAGES

The present version of IndexMatic cannot handle right-to-left or bi-directional texts. The script is mainly designed to process queries on Latin-based contents. (In a query you can insert any Unicode character though, using the syntax: `\uHHHH`.)

### 2. BOOK MANAGEMENT

When targeting an InDesign book, the script mutely opens the underlying documents in order to check various data and to scan chapters. This is usually transparent to the user and should not cause any issue! However, a few beta-testers reported unclarified bugs about book management. Under some circumstances IndexMatic cannot read the inner status of a chapter or might detect 'name collisions' between a book chapter and another existing document. This issue seems related to importing books or chapters from an older version of InDesign. In this case, consider to rebuild the book. Otherwise, try to manually open the chapters that you need to index before running the script.

### 3. QUERY ENGINE

(a) IndexMatic is not based on the InDesign GREP module, so it cannot interpret some GREP operators, or the POSIX shortcuts. Please refer to the **SPECIAL CHARACTERS IN REGULAR EXPRESSIONS** and **ADDITIONAL METACHARACTERS** sections to get a comprehensive view of the available operators.

(b) A query KEY is limited to 172 characters.

## VERSION HISTORY

### 2.025 October 8, 2011

MAJOR UPDATE. Fixed issues with regex queries using complementary classes in the form `[^xxx]`. Fixed issues with text-style-range and threaded frames 'segmentation'. Threaded footnote location is now correctly resolved. Improved algorithm in parsing 'cross-page' contents. The Generic Space now recognizes Forced Line Break. The max. key length is now: 172. Added metacharacters: `\w, \l, \L, \m, \M, \p{...}`. Added the Entry Separator feature. Added comment and cross-reference syntax (`// topic1 => See topic2`). Added the "Extends alphabet to underscore" option. New Query Editor with icon buttons (load/save, increase/decrease type size). Added the Hit Report feature.

### 2.024 July 23, 2011

IMPORTANT UPDATE. Solved Enter Key issues in the Query Editor (Mac). Fixed a serious bug related to capturing words in Automatic Mode or using `\w` in regular expressions. Added the Sub Sort feature (which applies specific order to subtopics).

### 2.022 July 6, 2011

MINOR UPDATE. Refined the Generic Space algorithm. Fixed some typos in UI dialog popups.

### 2.021 July 1, 2011

ORIGINAL PUBLIC RELEASE.

Further details, tips and tricks about IndexMatic<sup>2</sup>:

<http://www.indiscripts.com/category/projects/IndexMatic>

# IndexMatic<sup>2</sup> Manual

© Indiscripts.com | October 2011 | v. 2.025



Index

32

- ALPHABET** 8-9, 11-13, 21-22, 25, 31
  - *CHARACTERS*
  - language 2, 21-22, 25, 31
- ANCHORED OBJECTS** 6, 30
- BOOK (INDESIGN)** 2, 4-5, 7, 12, 31
- CAPTURE** → *MATCH*
- CASE** 7, 11, 16, 19, 21-24, 27-28, 30-31
- CHARACTERS**
  - apostrophe 9
  - ASCII 9
  - case → *CASE*
  - dash 22, 26
  - diacritics 9
  - digits 9, 17, 21-22
  - discretionary ~ 21
  - letters 8-9, 11, 21-22
  - meta~ → *REGEX*
  - non-breaking ~ 9, 11, 22
  - space ~ → *SPACES*
  - symbols 16, 18-19, 21-22
  - tab 5, 11, 21-23, 26
- CONTEXT** 6, 8, 20-21, 30
- CROSS-REFERENCE** → *QUERY*
- DASH** → *CHARACTERS*
- FILTER** 2, 6-7, 12, 29
- FOOTNOTES** 6, 30-31
- HITS** 2-3, 10-12, 21, 27, 29-30
- INDEX** 1-15, 18-31
  - ~ entries 19, 23-28, 31
  - export 27
  - formatting 1-3, 7, 11, 18, 23-24, 26, 28
  - page numbers → *PAGES*
  - separators 22-23, 26, 31
  - sorting 12, 17, 21, 23, 25-31
- LAYER** 4-5
- MATCH**
  - capture 9-13, 19, 21-22, 24, 29-30
  - forms, variants 3, 15, 19, 21, 31
  - placeholders 11, 15-16, 19, 24, 27
  - *QUERY*
- METACHARACTERS** → *REGEX*
- PAGE RANK** 2-3, 10, 12-13, 17, 27, 29
- PAGES**
  - 2-4, 8, 10, 12-13, 17, 20-21, 23-24, 26-31
  - 27
  - location
  - ~ number 2-4, 21, 23, 26, 29
  - range 4, 9, 26, 30
  - tolerance 26
- PARAGRAPH** 7, 12, 21-22, 28, 30
- PARENTHESIS** → *REGEX*
- QUERY**
  - comment 2-3, 8-22, 24-25, 27, 29-31
  - 14, 20, 31
  - cross-reference 14, 20, 31
  - fake term 20
  - flags 17
  - interpreter 15, 18, 20
  - key 15-21, 31
  - operators 2, 15-16, 18-21, 24, 31
  - pattern → *REGEX*
  - placeholders 19-21, 24
- term** 1, 15, 18-21, 24-25, 27, 29
- token** 16, 18-19, 24
- topic / subtopic** 2, 13-15, 18-20, 23-27, 31
- REGEX**
  - 2-3, 8, 12, 16, 19-22, 24, 31
  - character class 21-22, 31
  - dot 21, 26
  - 2, 11, 22, 31
  - GREP → *MATCH*
  - match → *MATCH*
  - metacharacters 9, 16, 21-22, 31
  - parenthesis 19, 21
  - pattern 9, 14, 16, 18-21, 24
  - Unicode 9, 11, 21-22, 31
- REGULAR EXPRESSIONS** → *REGEX*
- REWRITING** → *QUERY* | operators
- SCOPE** 3-4, 6-8, 10, 12, 29-30
- SORT** → *INDEX* | sorting
- SPACES**
  - 11-13, 15-17, 21-22, 31
  - generic ~ 11, 16-17, 21, 23, 25, 31
- STATS** 3, 29
- SUB SORT** → *INDEX* | sorting
- TAB** → *CHARACTERS*
- TABLES AND CELLS** 6, 9, 11, 21-22, 24, 30
- TEXT FRAME**
  - 4, 6, 28, 30
  - overset 30
  - story 4, 30
  - threaded ~ 30-31
- TOLERANCE** → *PAGES*
- XML** 2, 23, 27