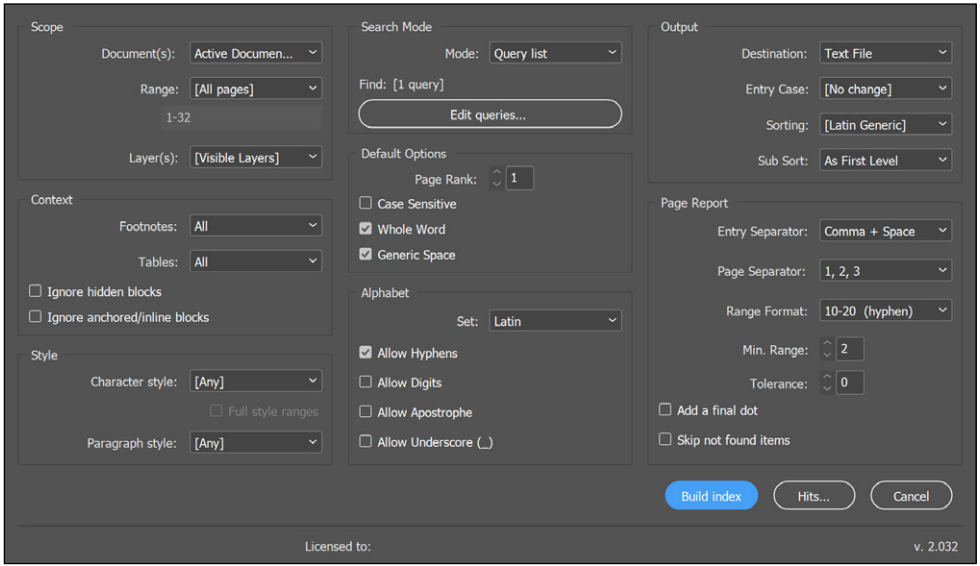
IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

Getting started



Main dialog of IndexMatic² (2.032, Windows UI).

1 INSTALLING & LAUNCHING INDEXMATIC²

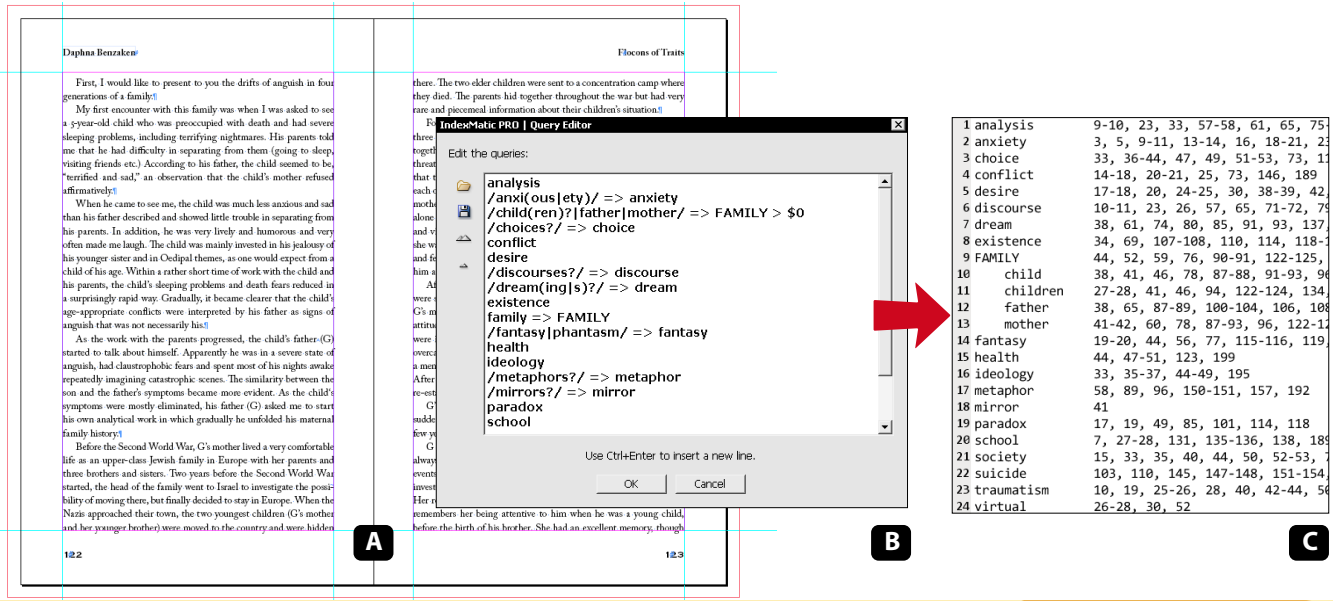
1. Download IndexMatic² from <indiscripts.com>.
   [PRO Version] Retrieve the package from your private link and save it on your hard disk.
   [TRY Version] Right-click IndexMatic2Try.zip and choose Save Target / Link As.
2. Extract IndexMatic2Pro.jsx (or ...Try.jsx) from the zip file, then place the 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.

- IndexMatic² supports CS4, CS5, CS6, CC 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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

Getting started



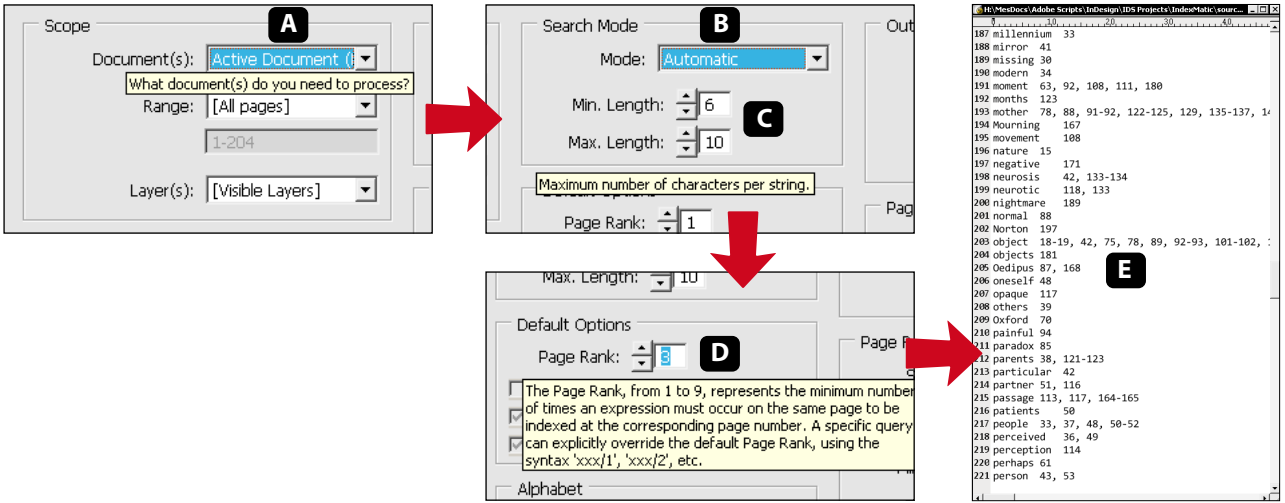
2 QUICK OVERVIEW

1. IndexMatic² 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. IndexMatic² 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).

- IndexMatic² sidelines the native InDesign 'Index' feature. The script never alter existing document(s): it just 'scans' the contents.
- Compared to other search tools, IndexMatic² usually provides excellent performance in that it does not employ the GREP engine at all.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



3 INDEXING YOUR FIRST DOCUMENT IN 15 SECONDS

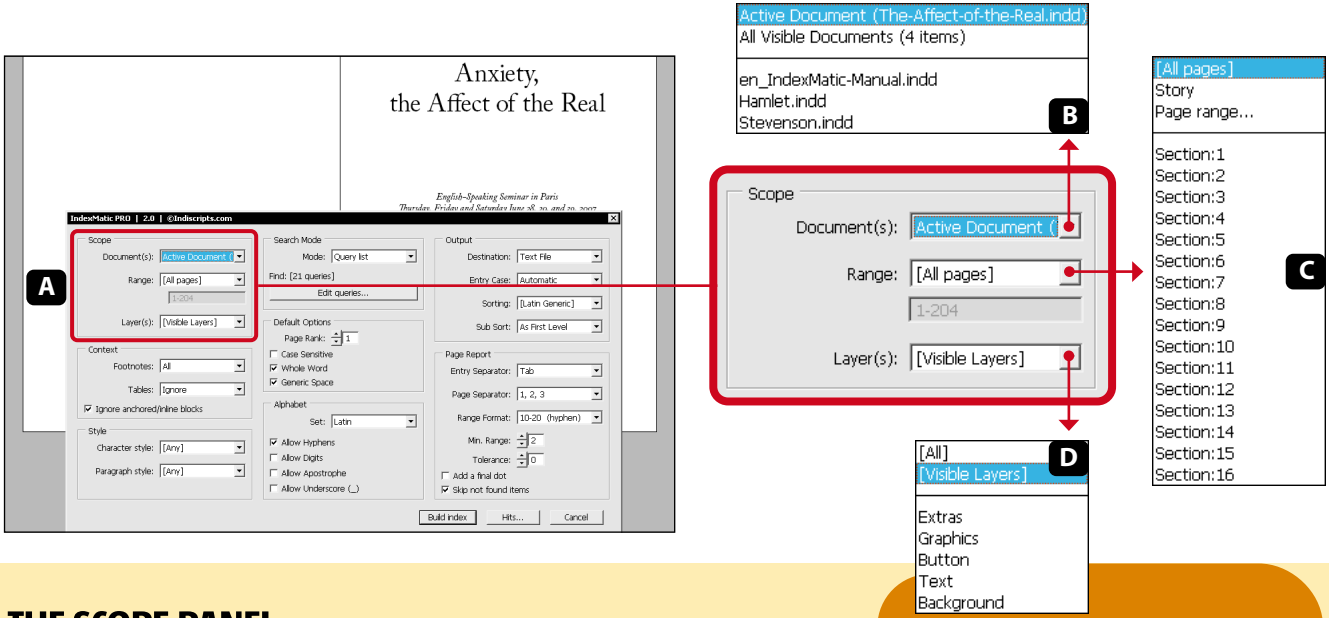
The simplest way to test IndexMatic² is to use the **Automatic** search mode (default). This hides the details of the query system and allows to quickly produce an index from scratch, without any explicit set of queries.

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.)

IndexMatic2 Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



4 THE SCOPE PANEL

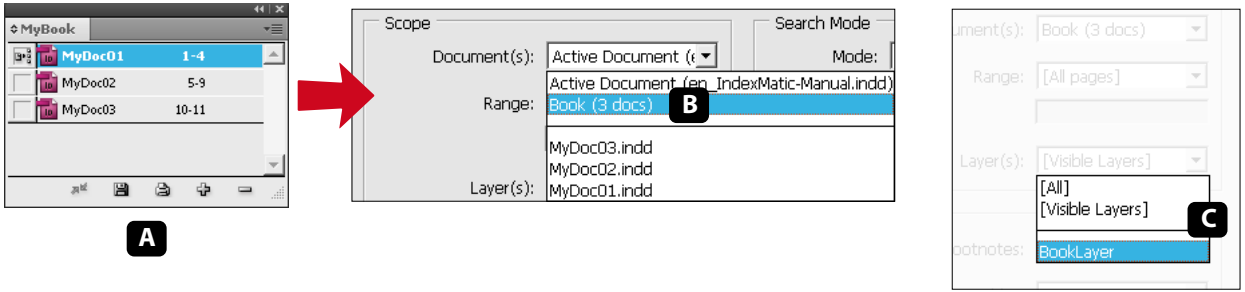
The **Scope** panel **A** indicates which document(s) to explore and defines the work area for indexing.

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)*

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

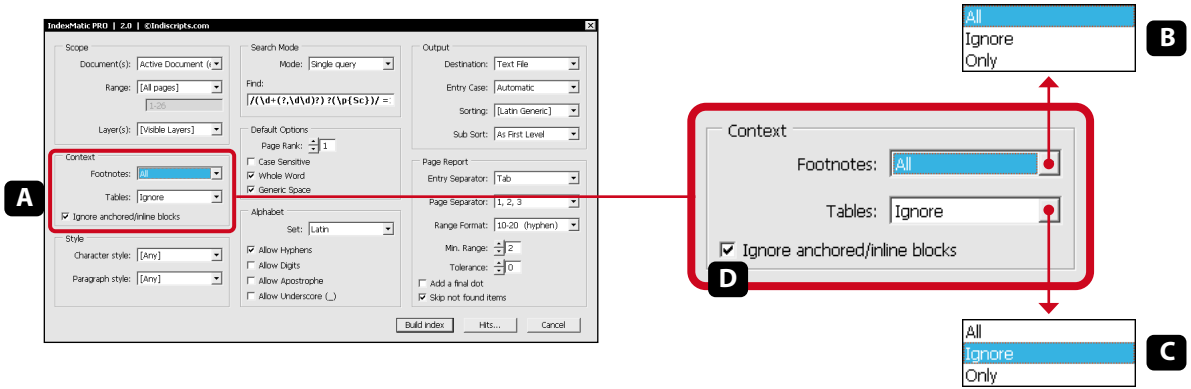


5 INDEXING A BOOK

Indexing an **entire book** with IndexMatic² 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 IndexMatic² 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.')

- IndexMatic² 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.



6 REFINING THE 'CONTEXT'

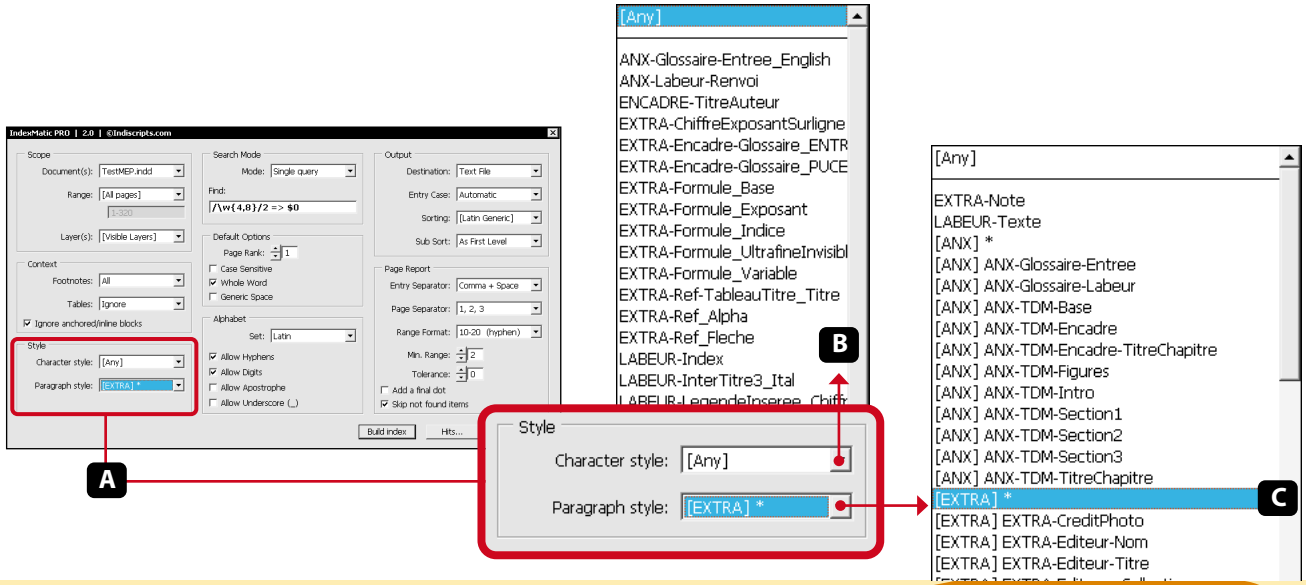
The **Context** panel **A** provides three additional filters over the Scope settings:

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 IndexMatic² 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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



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 IndexMatic² to selectively explore style-tagged contents within the current scope:

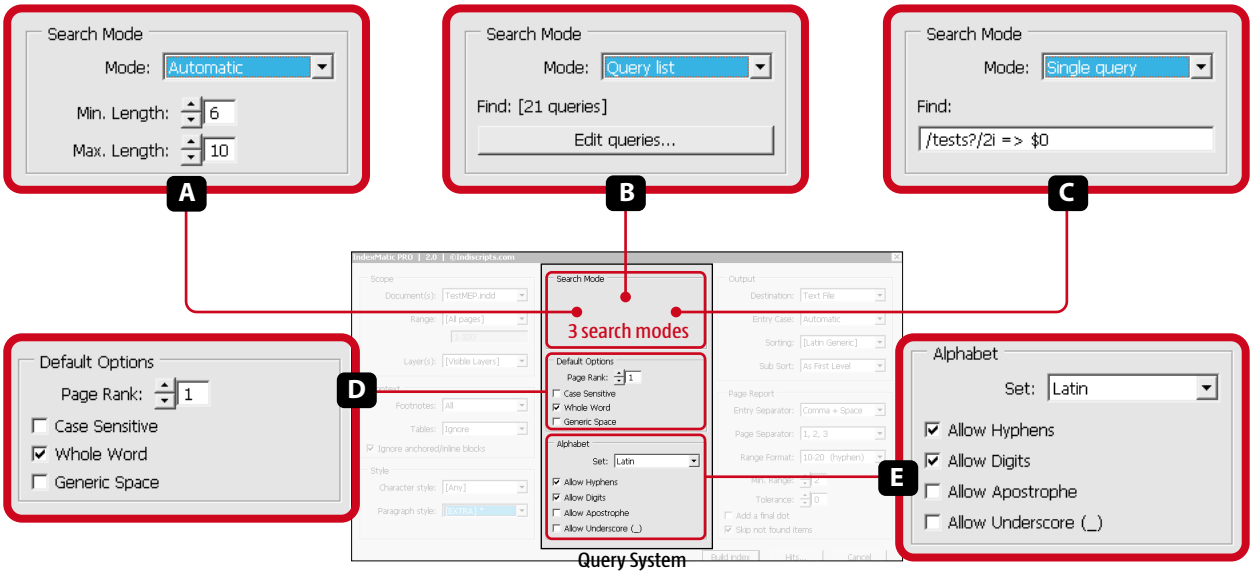
1. **Character style** **B** allows you to restrict the search field to a specific character style or style group.
2. **Paragraph style** **C** allows you to restrict the search field to a specific paragraph style or style group.

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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



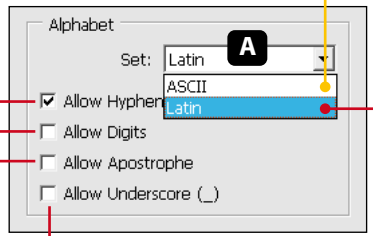
8 ABOUT THE QUERY SYSTEM

Once you have defined which parts of the document(s) you want to inspect *(see previous pages)*, you'll tell IndexMatic² 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, IndexMatic² 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.

Search Options



B Extends the alphabet to HYPHENS (i.e. Unicode HYPHEN-MINUS, HYPHEN and NON-BREAKING HYPHEN.) Note that IndexMatic does not distinguish between these three characters, and that it removes any soft hyphen from the text stream. Sample match: `lily-of-the-valley`

C Extends the alphabet to DIGITS (0-9). Useful if you want alphanumeric expressions to be regarded as whole words. Sample match: `IndexMatic2`

D Extends the alphabet to the APOSTROPHE (i.e. ASCII apostrophe and RIGHT SINGLE QUOTATION MARK). Sample match: `don't`

E Extends the alphabet to the UNDERSCORE (U+005F). Sample match: `abc_def`

Main characters supported by the 'Latin' set *(the table does not display all the glyphs)*. The formal range is: `U+0041-U+005A; U+0061-U+007A; U+00C0-U+00D6; U+00D8-U+00F6; U+00F8-U+02AF; U+1E00-U+1EFF.`

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
À Á Â Ã Ä Å Æ Ç È É Ê Ë Ì Í Î Ï Ð Ñ Ò Ó Ô Õ Ö Ø Ù Ú Û Ü Ý Þ ß
à á â ã ä å æ ç è é ê ë ì í î ï ð ñ ò ó ô õ ö ø ù ú û ü ý þ ÿ
Ā ā Ă ă Ą ą Ć ć Ĉ ĉ Ċ ċ Č č Ď ď Đ đ Ē ē Ĕ ĕ Ė ė Ę ę Ě ě Ĝ ĝ Ğ ğ Ġ
ġ Ģ ģ Ĥ ĥ Ħ ħ Ĩ ĩ Ī ī Ĭ ĭ Į į İ IJ ij Ĵ ĵ Ķ ķ ĸ Ĺ ĺ Ļ ļ Ľ ľ Ŀ ŀ Ł ł Ń ń Ņ ņ Ň ň ŉ
Ŋ ŋ Ō ō Ŏ ŏ Ő ő Œ œ Ŕ ŕ Ŗ ŗ Ř ř Ś ś Ŝ ŝ Ş ş Š š Ţ ţ Ť ť Ŧ ŧ Ũ ũ Ū ū
Ŭ ŭ Ů ů Ű ű Ų ų Ŵ ŵ Ŷ ŷ Ÿ Ź ź Ż ż Ž ž ſ ƀ Ɓ Ƃ ƃ Ƅ ƅ Ɔ Ƈ ƈ Ɖ Ɗ Ƌ ƌ ƍ
Ǝ Ə Ɛ Ƒ ƒ Ɠ Ɣ ƕ Ɩ Ɨ Ƙ ƙ ƚ ƛ Ɯ Ɲ ƞ Ɵ Ơ ơ Ƣ ƣ Ƥ ƥ Ʀ Ƨ ƨ Ʃ ƪ ƫ Ƭ ƭ Ʈ Ư
ư Ʊ Ʋ Ƴ ƴ Ƶ ƶ Ʒ Ǯ ǯ ȝ 2 5 ƽ ƾ ƿ ǀ ǁ ǂ ǃ DŽ Dž dž LJ Lj lj NJ Nj nj Ǎ ǎ Ǐ ǐ
Ǒ ǒ Ǔ ǔ Ǖ ǖ Ǘ ǘ Ǚ ǚ Ǜ ǜ ǝ Ǟ ǟ Ǡ ǡ Ǣ ǣ Ǥ ǥ Ǧ ǧ Ǩ ǩ Ǫ ǫ Ǭ ǭ DZ
Dz dz Ǵ ǵ Ƕ Ƿ Ǹ ǹ Ǻ ǻ Ȁ ȁ Ȃ ȃ Ȅ ȅ Ȇ ȇ Ȉ ȉ Ȋ ȋ Ȍ ȍ Ȏ ȏ Ȑ ȑ Ȓ ȓ Ȕ ȕ
Ą ą Ḃ ḃ Ḅ ḅ Ḇ ḇ Ḉ ḉ Ḋ ḋ Ḍ ḍ Ḏ ḏ Ḑ ḑ Ḓ ḓ Ḕ ḕ Ḗ ḗ Ḙ ḙ Ḛ ḛ Ḝ ḝ
Ḟ ḟ Ḡ ḡ Ḥ ḥ Ḧ ḧ Ḩ ḩ Ḫ ḫ Ḭ ḭ Ḯ ḯ Ḱ ḱ Ḳ ḳ Ḵ ḵ Ḷ ḷ Ḹ ḹ Ḻ ḻ Ḽ ḽ Ḿ ḿ
Ṁ ṁ Ṃ ṃ Ṅ ṅ Ṇ ṇ Ṉ ṉ Ṋ ṋ Ṍ ṍ Ṏ ṏ Ṑ ṑ Ṓ ṓ Ṕ ṕ Ṗ ṗ Ṙ ṙ Ṛ ṛ
Ṝ ṝ Ṟ ṟ Ṡ ṡ Ṣ ṣ Ṥ ṥ Ṧ ṧ Ṩ ṩ Ṫ ṫ Ṭ ṭ Ṯ ṯ Ṱ ṱ Ṳ ṳ Ṵ ṵ Ṷ ṷ Ṹ ṹ Ṻ ṻ Ṽ ṽ
Ṿ ṿ Ẁ ẁ Ẃ ẃ Ẅ ẅ Ẇ ẇ Ẉ ẉ Ẋ ẋ Ẍ ẍ Ẏ ẏ Ẑ ẑ Ẓ ẓ Ẕ ẕ ẖ ẗ ẘ ẙ ẚ
Ạ ạ Ả ả Ấ ấ Ầ ầ Ẩ ẩ Ẫ ẫ Ậ ậ Ắ ắ Ằ ằ Ẳ ẳ Ẵ ẵ Ặ ặ Ẹ ẹ Ẻ ẻ Ẽ ẽ
Ế ế Ề ề Ể ể Ễ ễ Ệ ệ Ỉ ỉ Ị ị Ọ ọ Ỏ ỏ Ố ố Ồ ồ Ổ ổ Ỗ ỗ Ộ ộ Ớ ớ Ờ ờ
Ở ở Ỡ ỡ Ợ ợ Ụ ụ Ủ ủ Ứ ứ Ừ ừ Ử ử Ữ ữ Ự ự Ỳ ỳ Ỵ ỵ Ỷ ỷ Ỹ ỹ
ɐ ɑ ɒ ɔ ɕ ɖ ɗ ɘ ə ɚ ɛ ɜ ɝ ɞ ɟ ɠ ɡ ɢ ɣ ɤ ɥ ɦ ɧ ɨ ɪ ɬ ɫ ɭ ɯ ɰ ɱ ɲ ɳ ɴ ɵ ɶ ɷ ɸ ɹ ɻ ɽ
ɾ ɿ ʀ ʁ ʂ ʃ ʄ ʅ ʇ ʈ ʉ ʊ ʋ ʌ ʍ ʏ ʐ ʑ ʒ ʔ ʕ ʖ ʗ ʘ ʙ ʚ ʛ ʜ ʝ ʞ ʟ ʠ ʡ ʢ ʣ ʤ ʥ ʦ ʧ ʨ

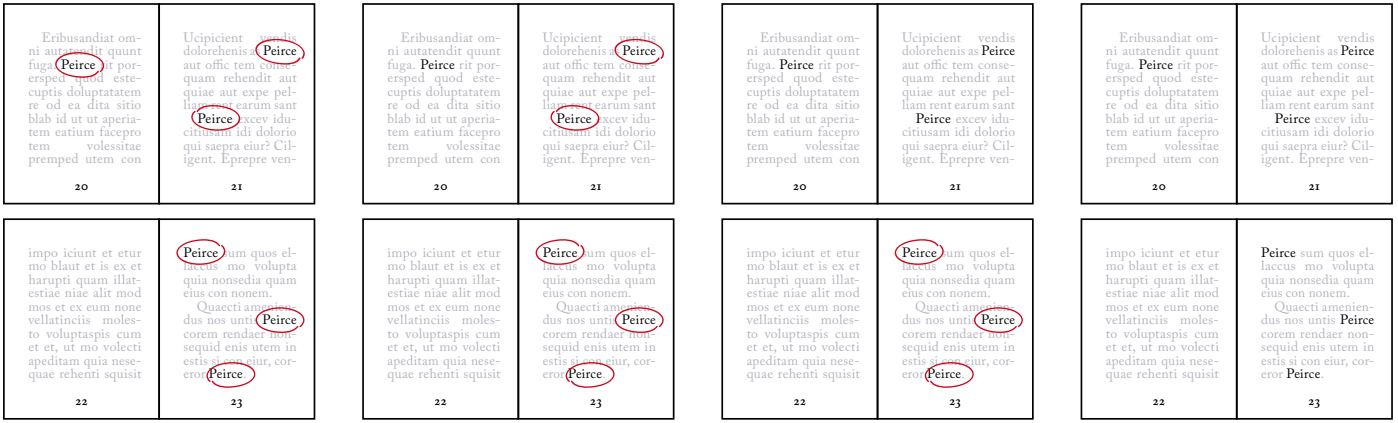
9 SETTING YOUR ALPHABET

The **Alphabet** panel is one of the most important features of IndexMatic² in that it affects the meaning of many search patterns and interacts with various matching options. An alphabet is a **limited set of characters** that are allowed to appear in a word. This directly effects the **Automatic** search mode and the **Whole Word** setting.

1. The **Set** field **A** offers two options: **ASCII** or **Latin**. 'ASCII' only contains the 26 basic letters **without diacritics**. 'Latin' includes **additional Unicode sets**: Latin-1 Suppl., Latin Extended-A/-B, IPA, and Latin Ext. Additional.
2. The **Alphabet** panel also provides four checkboxes that let you add, if needed, some non-literal characters: **Allow Hyphens** **B** (checked by default), **Allow Digits** **C**, **Allow Apostrophe** **D**, and **Allow Underscore** **E**.

- The selected alphabet defines what a word letter is, so every character that does not belong to the set is assumed to be 'outside' of any word.
- IndexMatic² only supports left-to-right Latin-based scripts. →31
- In regex-based queries, the following metacharacters →21 are automatically adjusted to the selected aphabet: `\w \W \l \L \m \M`

Search Options



10 UNDERSTANDING THE 'PAGE RANK'

In IndexMatic2, 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."

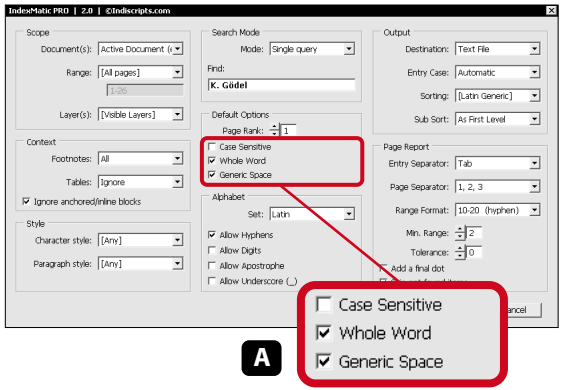
1. By setting a page rank greater than 1—say 3—in the **Default Options** panel, you force the query system to ignore any match that does not occur at least 3 times in a given page.
2. Increasing the page rank is particularly effective when using the **Automatic** search mode as it usually produces a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

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

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	ˇ
Hair Space	~|	U+200A	⁚
Medium Mathematical Space		U+205F	¬

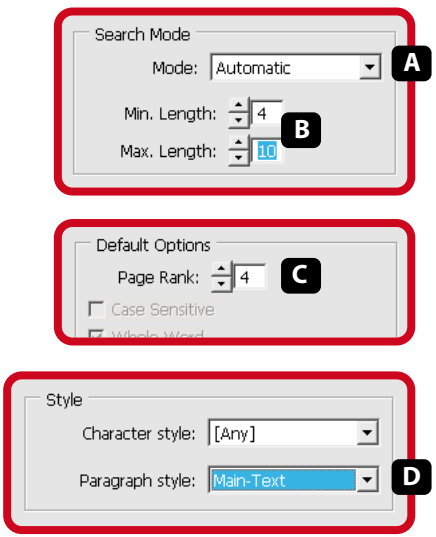
11 MATCHING OPTIONS

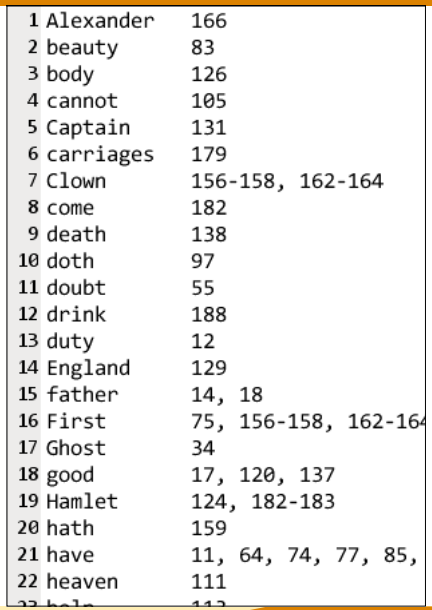
The **Default Options** panel offers three options **A** that can either enforce or relax the matching constraints. Suppose we need to index the occurrences of `K. Gödel` (this could easily be done through a Single query):

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.

Search Options





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

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.

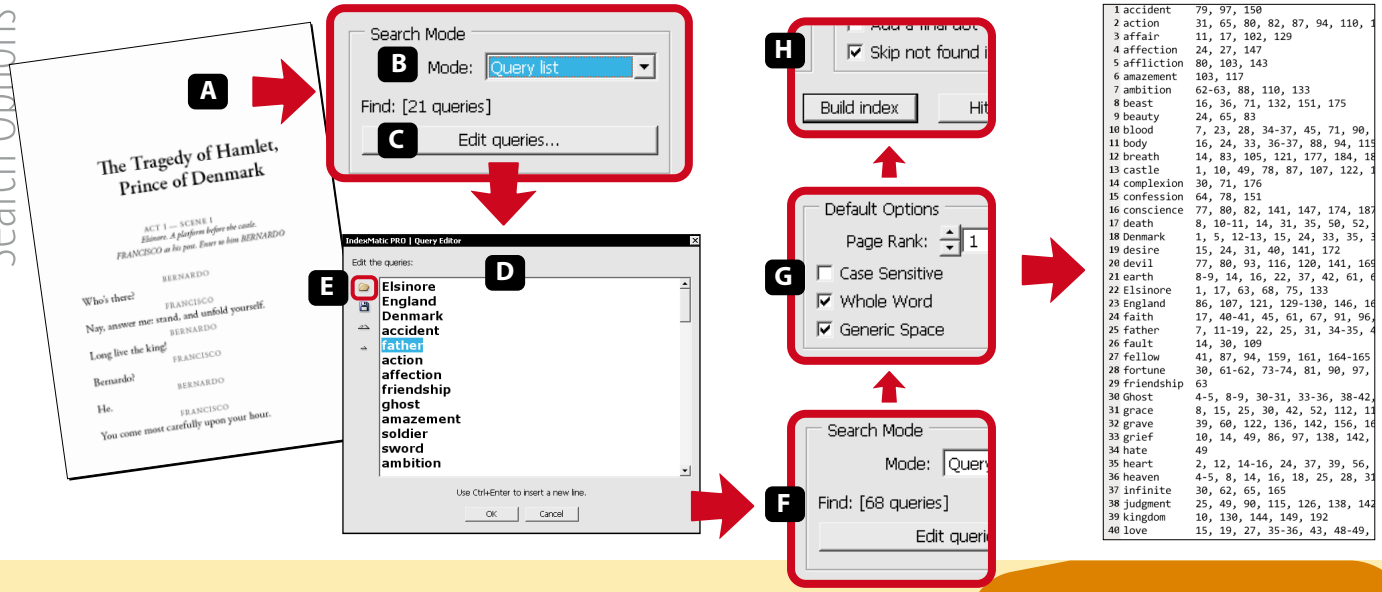
1. In the **Search Mode** dropdown list, select 'Automatic' **A**.
2. Adjust the **Min. Length** (>= 2) and **Max. Length** (<= 40) **B**. The query system will search in the scope any sequence of characters forming a 'whole word' according to the rules of the current **alphabet** →**9**.
3. If you are 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 →**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 →**19**, such as: `/.+/`

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

Search Options



13 TURNING A WORD LIST INTO AN INDEX

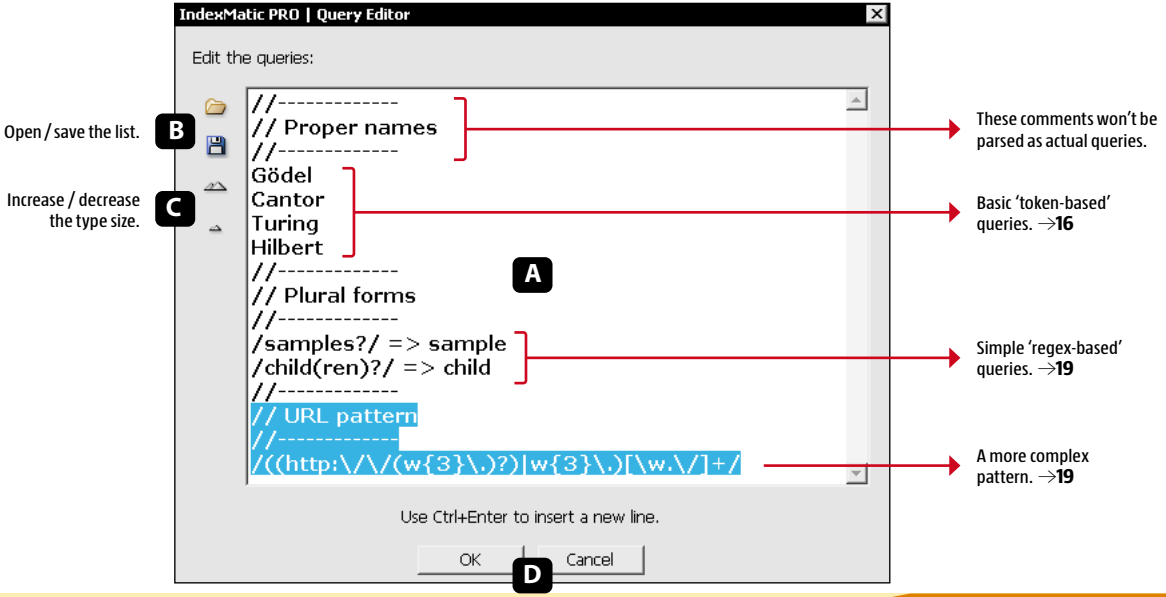
Suppose you've just finished laying out "Hamlet" A and the publisher wants you to carry out a subject index based on an handful of predefined topics—such as: *love*, *death*, *madness*, *Denmark*...

1. Run IndexMatic² 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: *bull***et**, *bull***etin**, *bull***ion**...
- 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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



14 USING THE QUERY EDITOR

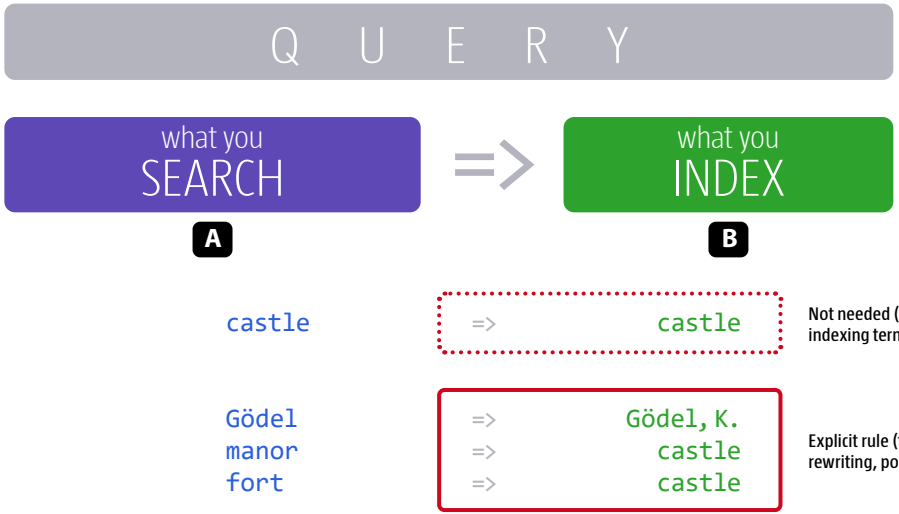
The **Query Editor** is a windowed interface available from the 'Query list' mode and the 'Edit queries...' button. It allows you to specify, edit, and manage a list of queries to be processed by the query engine →**15**.

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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



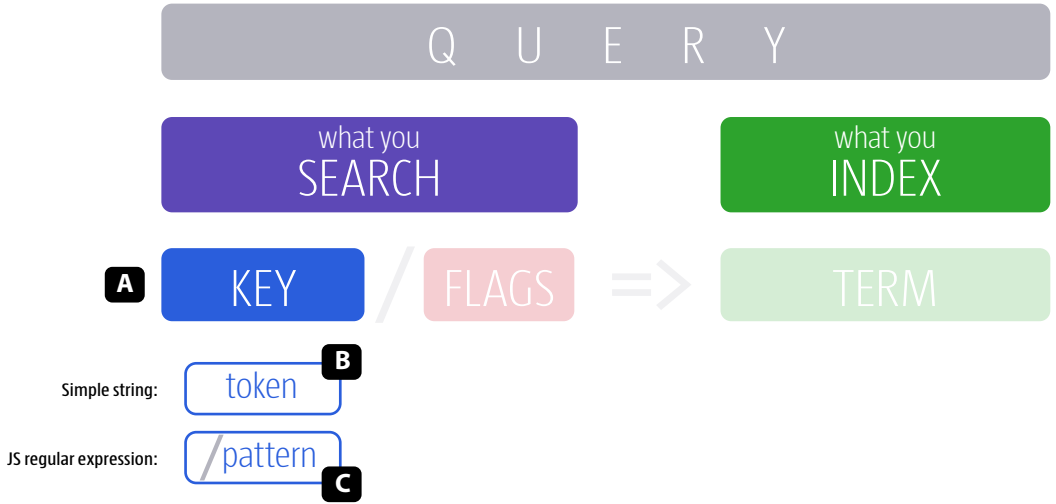
15 ABOUT THE QUERY SYNTAX

1. What makes IndexMatic² 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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



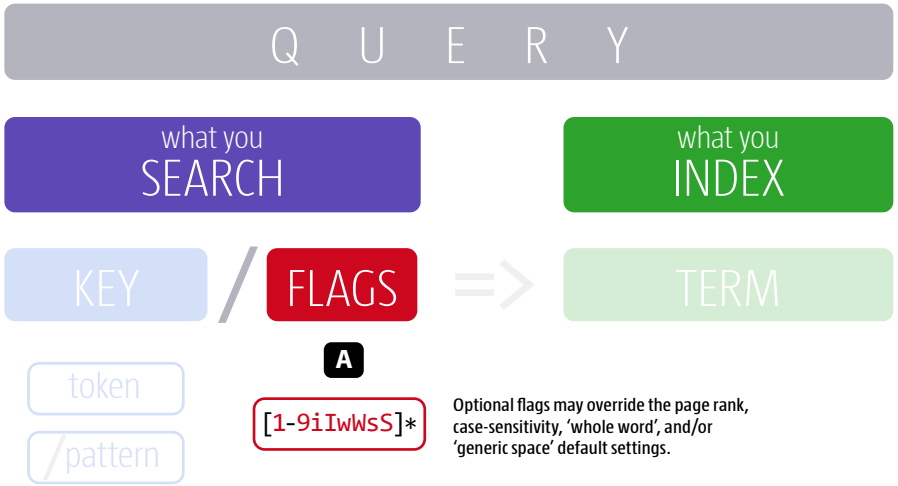
16 QUERY SYNTAX: THE 'KEY'

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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



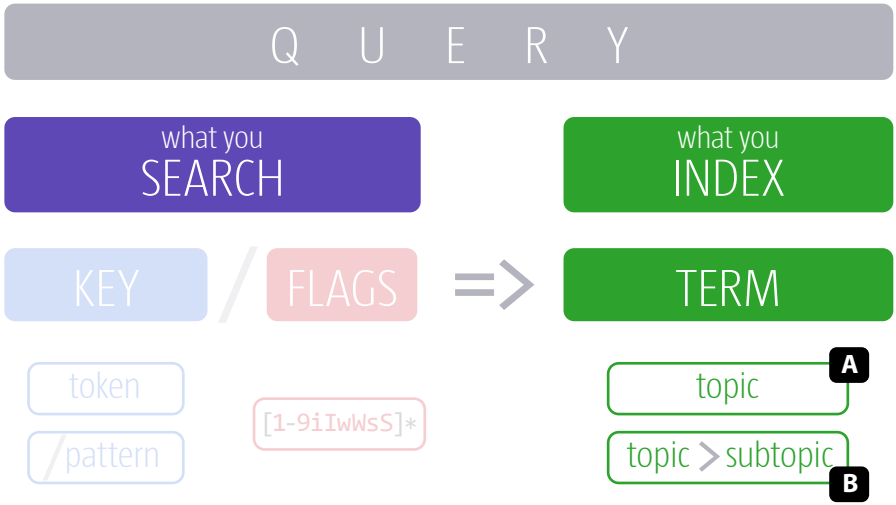
17 QUERY SYNTAX: THE 'FLAGS'

1. Optional **FLAGS** **A** 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.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



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 (**=>**).

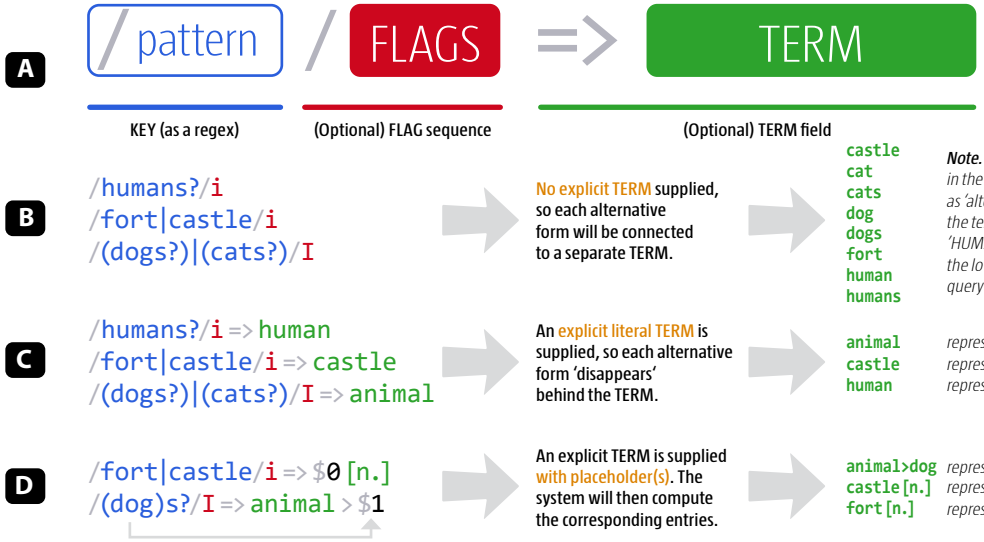
1. In its basic form, an explicit TERM is a string, such as `animal`. It then "takes the place" of the KEY and serves as a topic in the final index **A**. For example: `dog=>animal` tells the interpreter to connect the token `dog` to the topic `animal`. (Of course you can re-use a same TERM in different queries.)
2. The TERM field also supports a special operator, **>**, that lets you both create a subtopic and specify a parent topic **B**. 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). IndexMatic² 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^2 Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

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^2 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^{th} 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?**

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

A

B

```
// castle => See also fort.
// animal > dog => See canine.
```

20 EMULATING CROSS-REFERENCES

1. A special comment syntax is allowed to indicate a cross-reference **A**: `// TERM => 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 →**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 →**23** instead of a page location.
- Both the comment marker (//) and the => operator are necessary to create a valid cross-reference. If the => operator is omitted, the 'query' is parsed as a regular comment→**14**.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

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

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

(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).*

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

A

■ 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: \$.

21 SPECIAL CHARACTERS IN REGULAR EXPRESSIONS

1. IndexMatic² 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*
2. In addition, IndexMatic² 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.)*

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

A Unicode Properties

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

Property	Use	Property	Use
\p{Ll}	Lowercase letter. E.g.: é, α, ā, œ, ç...	\p{Ps}	Open punctuation. E.g.: (, [, {...
\p{Lu}	Uppercase letter. E.g.: A, É, Đ, Ω, NJ...	\p{Pe}	Close punctuation. E.g.:),], }...
\p{Lt}	Titlecase letter. E.g.: Dž, Lj, Nj...	\p{Pi}	Initial punctuation. E.g.: «, ‘, "...
\p{Lm}	Modifier letter. E.g.: ʰ, ˣ, ʕ...	\p{Pf}	Final punctuation. E.g.: », ’, "...
\p{Lo}	Letter without case. E.g.: ƻ, ǁ, ಹ...	\p{Pc}	Connector punctuation. E.g.: _, ‿, ⁀ ...
\p{M}	Any mark.	\p{Po}	Other punctuation. E.g.: !, ;, #...
\p{Mn}	Non-spacing mark.	\p{S}	Any symbol.
\p{Mc}	Spacing combining mark.	\p{Sm}	Math symbol. E.g.: +, <, ↔...
\p{Me}	Enclosing mark.	\p{Sc}	Currency symbol. E.g.: $, €, £, ৳...
\p{N}	Any number.	\p{Sk}	Modifier symbol. E.g.: ^, ¯, ¸, ˜...
\p{Nd}	Decimal digit. E.g.: 1, 2, ٤...	\p{So}	Other symbol. E.g.: ¦, §, ©...
\p{Nl}	Letter number. E.g.: viii, ⑯, ↻...	\p{Z}	Any separator.
\p{No}	Other number. E.g.: ², ⅋, ⅏...	\p{Zs}	Space separator.
\p{P}	Any punctuation.	\p{C}	Any other character.
\p{Pd}	Dash punctuation. E.g.: -, —, ~...		

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

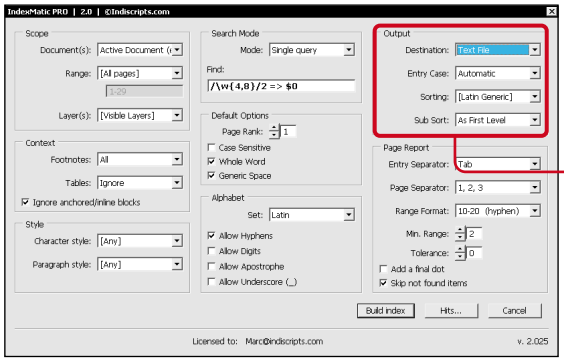
B Additional GREP shortcuts

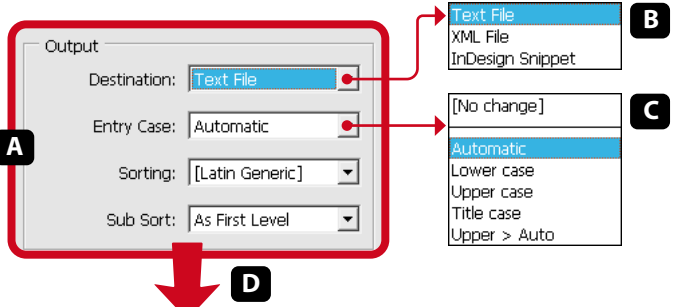
Op.	Character	
~8	Bullet	•
~e	Ellipsis	…
~7	Paragraph Symbol	¶
~6	Section Symbol	§
~2	Copyright	©
~r	Registered Trademark	®
~d	Trademark Symb.	™
~_	Em Dash	—
~=	En Dash	–
~{	Double Left Quot.	“
~}	Double Right Quot.	”
~[	Single Left Quot.	‘
~]	Single Right Quot.	’
~"	Straight Dble Quot.	"
~'	Straight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

22 ADDITIONAL METACHARACTERS

1. Since v. 2.025, IndexMatic² 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}**
2. The query engine also supports a special subset of GREP metacharacters **B**.
3. 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 IndexMatic² is not identical to InDesign GREP syntax—although there are considerable similarities between the two languages.**






```
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,
```

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.

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.

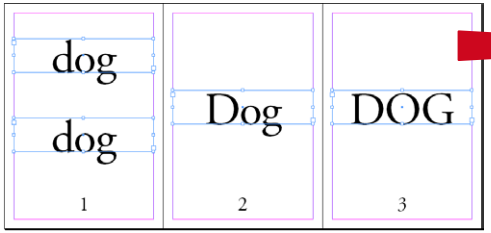
1. In the **Destination** list, select 'Text File' (default) **B** to get the resulting index in an editable **plain text file**. The other options are 'XML File' →**27** and 'InDesign Snippet' (CS4/CS5 only)→**28**.
2. In the **Entry Case** list, select 'Automatic' (default) **C** to let IndexMatic² 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 desktop.)
- The **TRY version** only allows 'Text File' destination.

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

Output



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 / ENTRY CASE	dog/i **A**	/do\w/i=>$0 **B**	/Dog/I **C**	dog/i=>Animal **D**	/dog/i=>animal>$0 **E**
[No Change]	dog 1-3	dog 1 Dog 2 DOG 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

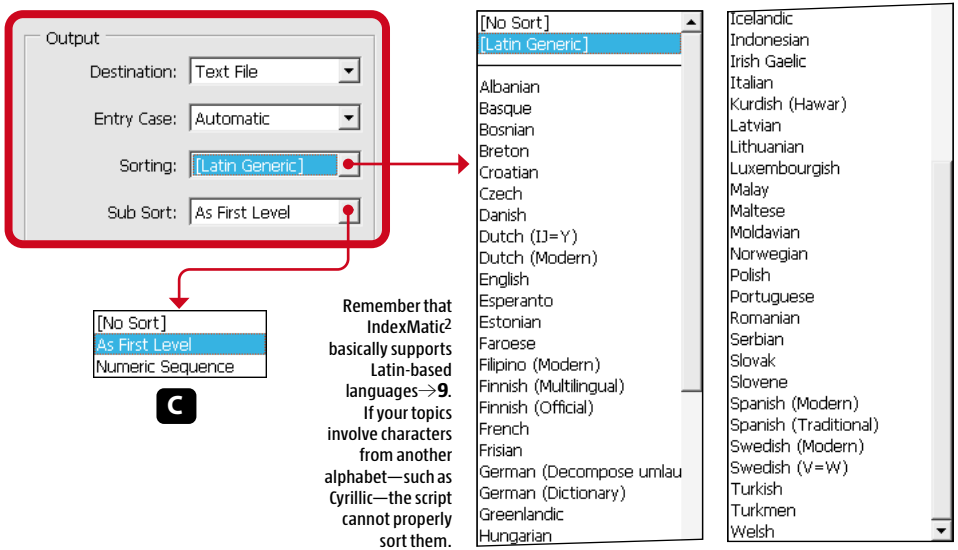
24 UNDERSTANDING THE 'ENTRY CASE' OPTIONS

1. 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., IndexMatic² internally creates two templates, {dog} and {dogs}, and maintain a counter for case variants.
2. 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, IndexMatic² 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**

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*



Remember that IndexMatic² basically supports Latin-based languages→**9**. If your topics involve characters from another alphabet—such as Cyrillic—the script cannot properly sort them.

EXAMPLE: Ordering Norwegian topics.

Sorting: [Latin Generic] **A**	Sorting: Norwegian **B**
åbner	altså
ære	bryst
altså	brænder
brænder	brød
brød	fjord
bryst	jeg
fjord	ordet
jeg	vild
øje	ære
ordet	øje
vild	åbner

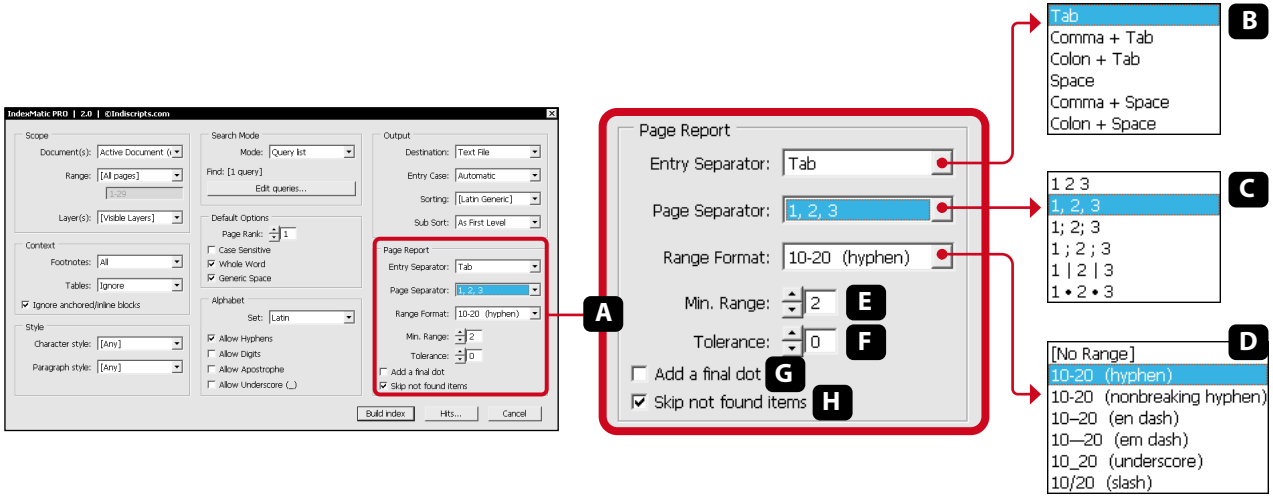
25 ORDERING THE FINAL INDEX ENTRIES

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 IndexMatic² 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.

Output



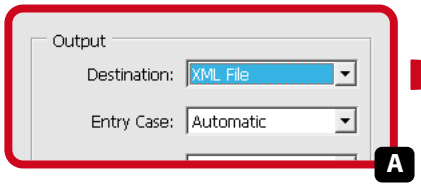
26 FORMATTING THE PAGE NUMBERS

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

1. 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**).
2. 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 box **G** to have each line ending with a dot point.
3. 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.)

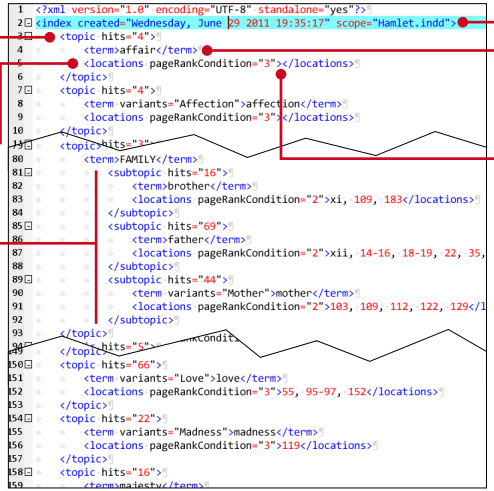
Output



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)



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).

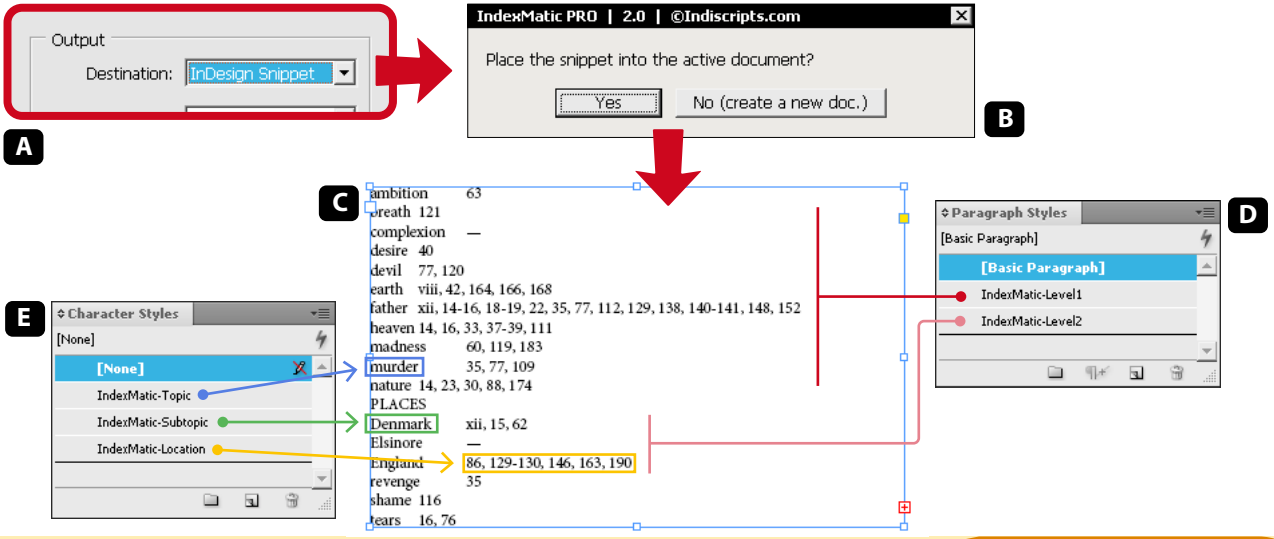
27 XML EXPORT

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.)

IndexMatic² Manual

© Indiscripts.com | October 2011 | v. 2.032 *(revision Feb., 2018)*

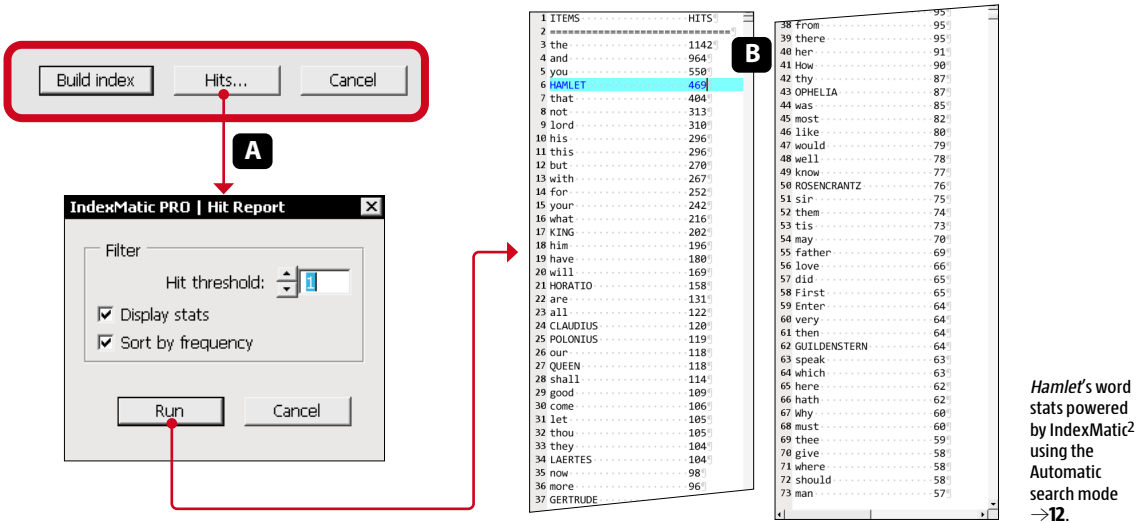


28 PLACING THE INDEX AS A 'SNIPPET' (CS4/CS5)

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, IndexMatic² 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.

Output



Hamlet's word stats powered by IndexMatic² using the Automatic search mode →**12**.

29

GETTING MATCHES & STATS BEFORE INDEXING

Since v. 2.025, IndexMatic² 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, foonote 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.

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

2.032 **February 19, 2018**
IMPORTANT UPDATE. Added the 'Full style range' checkbox with the Character style list. This option makes explicit the ability to consider entire character style ranges when the Whole Word option is active. (In version 2.031 this behavior was forced.) + Minor fixes making the script more CC friendly.

[See detail on intermediate revisions of the script in your readme file.]

2.021 **July 1, 2011**
ORIGINAL PUBLIC RELEASE.

Index