

# Stemming Selector 4

## User Guide

## CONTENTS

<b>Quick Start</b>	<b>3</b>
<b>Installation &amp; Registration</b>	<b>4</b>
<b>Command-Line Options</b>	<b>5</b>
<b>Tables</b>	<b>6</b>
<b>License</b>	<b>7</b>

## Quick Start

Stemming Selector 4 is for use with dtSearch Desktop or dtSearch Network to provide language specific stemming. Because dtSearch Desktop uses the stemming rules at search time (known as Query Conflation) there is no need to re-index for each language and no need to have language specific indexes.

Language specific stemming is essential to obtain the high recall demanded by professionals in the legal, law enforcement, intelligence, and other serious research fields for non-English data. \*

Stemming Selector 4 is distributed on its own for use in conjunction with dtSearch Desktop or dtSearch Network or as part of a Language Extension Pack (LEP500 series) for use by end-users or developers.

1. From the **Settings** menu, choose **Target Path...** (see Fig 1) and browse to the dtSearch stemming.dat file; this will normally be found under the c:\Program Files (x86)\dtSearch\bin\ folder (Windows 7) or c:\Program Files\dtSearch\bin\ folder (on Windows XP).



Fig 1

2. Select a language from the drop-down list, dtSearch will use the stemming language that you have selected to expand your search query, there is no need to update the index. You can change the stemming language at any time with dtSearch running or not.

3. In the dtSearch Desktop search dialog make sure Stemming is selected. You will notice that with stemming selected you will normally get more hits in your search results than with it unselected.

4. You can either close the Stemming Selector or minimise it to the Windows task bar, your settings are automatically saved. If you are frequently changing the stemming language you may find it more convenient to choose the **Always on top** option from the **Settings** menu. (You will need to unselect **Always on top** if you need to access the Help|About dialog).

The languages that are displayed in the drop-down list depend on the product license purchased as below.

**SLS500 or LEP500 - All languages listed in Tables A, B and C (see page 6)**

**SLS502 or LEP502 - All languages listed in Table A**

**SLS503 or LEP503 - All languages listed in Table B**

### **LEP500, LEP502, LEP503**

If you have purchased a Language Extension Pack you will also have test files and stop word lists for all the languages in Tables A, B and C. You can find details of how these are used in the LEP500 Guide. The Language Pack licenses also include ability to generate multiple stemming language files in the same format as the previous LEP400 series, these can be distributed with an application to the extent of your license. See **Command-line Options** on page 5.

\* User Thesaurus Plus is another add-on product that can further enhance recall and supplement the stemming feature; it is supplied with a number of thesauri including irregular verbs in English, German, Dutch, Spanish and Italian, and English irregular nouns. User Thesaurus Plus is included with LEP500, LEP502 and LEP503 language packs for distribution with an application to the extent of your license, or can also be licensed separately for end-user use only.

## Installation & Registration

You must have dtSearch Desktop or dtSearch Network installed on your machine before trying to install the Stemming Selector 4. This version of the Stemming Selector does not require the install of LEP400 or LEP500, it can be used on its own in conjunction with dtSearch Desktop/Network to change the stemming language at any time.

### Installation

Download (Save...) the file SLS500\_Setup.zip to the machine where you want to install it, right-click on the file and choose Extract All... , choose a path where you want to save the files (e.g. your My Downloads folder). The files will be extracted to a folder, open the folder and double-click on setup.exe to run it, then follow the instructions. The installer will create a short-cut on your Windows Programs menu named **Stemming Selector** under the folder DTSUK.

### Evaluation Version

When first installed the program will run in evaluation mode. The evaluation version will run for 30-days, it will allow you to select all the languages shown in Tables A & B. After the 30-days has elapsed it will revert to English. You can enter a serial number at any time during or after the evaluation period.

### Registration

If you have purchased a license for LEP500, LEP502, LEP503, SLS500, SLS502 or SLS503 you will need to enter the serial number to unlock the languages and other options that your product license covers. From the **Help | About** menu press the **Add serial...** button (Fig 2). Enter the serial number and press OK. (If you have entered an invalid number the text boxes will turn red). We recommend that you register your purchase online at [www.dtsearch.co.uk](http://www.dtsearch.co.uk), from the Support menu choose Register product.

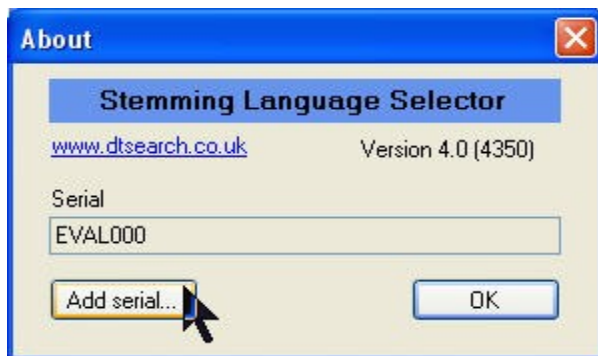


Fig 2

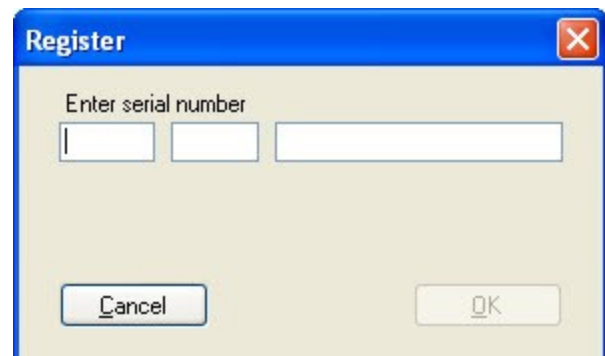


Fig 3

### Uninstalling

If you need to uninstall go to **Windows Settings | Control Panel | Add or Remove Programs** and click on StemSelect4 (or LEP500 if it was installed as part of a Language Extension pack).

## Command-Line Options

- <b>rusp</b>	Choose this to add the bi-lingual 'Russian Plus' stemming rules to the drop-down list. These are useful if you have a collection of documents that have mixed text in Russian (Cyrillic) and one of the Latin character based languages as shown in Table C.	(LEP500 version only)  The status strip will show R+ if this is applied.
- <b>exp</b>	<p>This option will export the stemming languages file that are selected in the drop-down list to the Target folder.</p> <p>1) Select the Target folder from the Settings menu (we recommend using the ...LEP500\Language\Stemming folder, select the stemming.dat file; if you use a different folder it will need to contain a file named stemming.dat).</p> <p>2) Select the stemming language rules that you require. A quick way to export them all is to select the first item in the drop-down list, then press the down-arrow key on your keyboard to reach the end of the list. The files will be exported in the ST_ <i>language</i>.dat file name format as used in previous versions of the Language Pack.</p>	(LEP500, LEP502, LEP503 versions only)  The status strip will show EXP if this is applied.

To run StemSelect4.exe with command line options , create a short-cut on your desktop then right-click on it and choose Properties; enter the command line in the Target window as shown below (Fig 4).

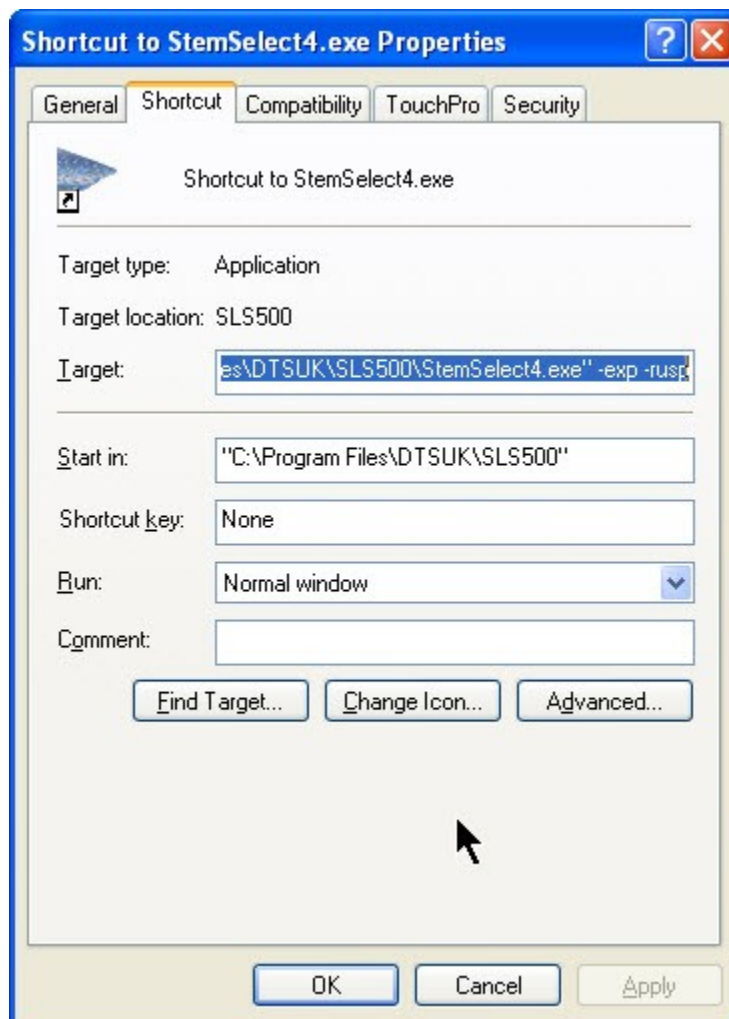


Fig 4. Short-cut Properties dialog

**Tables**

<b>TABLE A</b>
Danish
Dutch
English
Finnish
French + English
German
German + English
Italian
Norwegian
Portuguese
Spanish
Swedish

<b>TABLE B</b>
Belarusian
Bosnian
Bulgarian
Croatian
Czech
English
Estonian
Greek
Hungarian
Latvian
Lithuanian
Polish
Romanian
Russian
Serbian
Slovak
Slovenian
Turkish
Ukrainian
Uzbek

<b>TABLE C</b>
Russian + Czech
Russian + Estonian
Russian + Finnish
Russian + German
Russian + Greek
Russian + Hungarian
Russian + Latvian
Russian + Lithuanian
Russian + Polish
Russian + Slovak
Russian + Swedish

## **License Agreement for the Stemming Language Selector ("Licensed Software") for dtSearch.**

By installing this software you (the Licensee) agree to the terms of this License Agreement, if you do not agree then do not proceed with the installation. "Licensor" is dtSearch UK. Registered Office: 27 Shaftesbury Centre, Barlby Road, London W10 6BN. United Kingdom Tel: 0845 299 7307 Fax: +44 (0)207 900 6021 www.dtsearch.co.uk

### **License**

In the absence of a signed written agreement with dtSearch UK, or unless purchased as part of a Language Extension Pack LEP500 series license, this Software is licensed for use by a single individual. The single individual may use the software on up to two computers with dtSearch Desktop, so long as the dtSearch Desktop application will be used on both computers exclusively by that individual. This product is an end-user application only and may not be distributed with any other application that uses the dtSearch developer API. If wider distribution is need please upgrade to a Language Extension Pack license.

If purchased with a Language Extension Pack LEP500 series, the license supplied with the Language Extension Pack will apply, including any distribution with an application incorporating the dtSearch Text Retrieval Engine under a license issued by dtSearch Corp.

2.1 The Licensed Software is protected by copyright laws and other international treaties, as well as other intellectual property laws and treaties. The Licensed Software is licensed, not sold, according to the terms of this License Agreement.

2.2 Licensor grants, and Licensee hereby accepts, an irrevocable, non-exclusive, and world-wide license to use the Licensed Software.

2.3 Except as expressly provided herein, Licensee may not distribute, copy, reproduce, sub-license, sell or otherwise transfer the Licensed Software. Licensee may make backup copies of the Licensed Software strictly for Licensee's archival purposes.

2.4 Licensee may not modify, adapt, translate or create derivative works based on the Licensed Software, with the exception of those Re-distributable files identified as being Modifiable, and only to the extent specified in the documentation accompanying the Licensed Software. Any copies that Licensee is permitted to make pursuant to this License Agreement must contain the same copyright and other proprietary notices that appear on or in the Licensed Software.

2.5 Licensee may not assign or otherwise transfer its rights under this agreement without written permission from the Licensor, which permission shall not be unreasonably withheld, except in the case of a sale of substantially all assets of Licensee's company or the merger or acquisition of Licensee's company into a new company, in which case Licensee may transfer its rights under this agreement to the acquiring party.

### **3. Maintenance and Support**

3.1 Licensor agrees to make available to Licensee any maintenance releases, new and enhanced versions or upgrades of the Licensed Software for a period of one year.

3.2 Licensor agrees to provide technical support via email to a single point of contact designated by the Licensee for a period of one year from the date of this license.

### **4 Limitations on Warranty and Liability.**

4.1 The Licensed Software is provided AS IS. To the extent permitted by applicable law, ANY AND ALL OTHER REPRESENTATIONS AND WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NON-INFRINGEMENT, except as stated in section 4.4, below, ARE EXPRESSLY EXCLUDED AND DISCLAIMED. LICENSOR AND ITS SUPPLIERS DO NOT AND CANNOT WARRANT THE PERFORMANCE OR RESULTS LICENSEE

OR LICENSEE'S END-USERS MAY OBTAIN BY USING THE SOFTWARE.

4.2 IN NO EVENT SHALL LICENSOR BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES, including lost profits, lost savings, lost opportunities or other incidental or consequential damages arising out of the use of or inability to use the Licensed Software, even if Licensor has been advised of the possibility of such damages.

4.3 UNDER NO CIRCUMSTANCE MAY LICENSOR'S LIABILITY TO LICENSEE, UNDER ANY AND ALL PROVISIONS OF THIS AGREEMENT, EXCEED THE LICENSE FEE.

4.4 To the extent that Licensor remains legally liable to Licensee, such liability shall expire one year from the date of this Agreement.

4.5 Subject to the limitations on liability contained in sections 4.1, 4.2, 4.3, and 4.4 Licensor represents and warrants: that it owns the Licensed Software or has requisite authority to enter into this transaction; that the Licensed Software, when properly used as contemplated herein, will not infringe or misappropriate any copyright, trademark, or the trade secrets of any third persons; and that to the best of Licensor's knowledge, the Licensed Software does not infringe any patents of third persons.

#### **General Terms.**

5.1 The failure to immediately enforce any provisions, rights or remedies under this contract shall not constitute a waiver by the party failing to enforce such provision, even if the party failing to enforce such provisions, rights or remedies is aware of the other party's contractual breach.

5.2 This agreement shall be interpreted under English Law.

5.3 Either party may terminate this agreement upon written notice if the other party materially violates any provision of this agreement and fails to remedy such violation within twenty-eight (28) days of receipt of a second written notice thereof, if such violation has not already been remedied following twenty-eight (28) days from receipt of a first written notice thereof. Evidence of delivery of such written notice, by recorded delivery for instance, is required.

5.4 This agreement contains the entire agreement between the parties, superseding all previous agreements. This agreement may not be amended other than by a written agreement.

5.5 If any portion of this agreement is found to be invalid, the remainder shall continue in force.

dtSearch is a registered trade mark of dtSearch Corp. Inc.  
dtSearch UK is a trading name of ElectronArt Design Ltd.  
©Copyright 1995-2012 Electronart Design Ltd. All Rights Reserved.