SRXEditor User Guide

Copyright © 2008 - 2016 Maxprograms

Table of Contents

Introduction **1**
SRXEditor 1
Supported Platforms 1
Basic Concepts 1
Segmentation Rules eXchange (SRX) 1
Regular Expressions 2
User Interface **3**
SRXEditor GUI 3
Menus 3
File Menu 3
Options Menu 4
Help Menu 4
Common Tasks **5**
Create SRX File 5
Edit existing SRX File 5
Test Segmentation Rules 6

Introduction

SRXEditor

SRXEditor is a cross-platform editor of segmentation rules, designed to use Segmentation Rules eXchange (SRX) 2.0, the open XML-based standard published by LISA.

You can use SRXEditor to create new SRX files or to edit existing ones. It can also be used for testing segmentation rules to ensure that they break text as expected.

SRXEditor includes a sample file in SRX 2.0 format with a default set of segmentation rules supporting most standard cases. It also includes segmentation rules specific for these languages:

Catalan	German
Chinese	Italian
Czech	Japanese
Danish	Polish
Dutch	Portuguese
English	Spanish
Finnish	Swedish
French	Thai

Supported Platforms

- Microsoft Windows 7, 8.x and 10
- Mac OS X 10.8/10.9/10.10/10.11 (64 bit)
- Linux with GNOME desktop (64 bit)

Basic Concepts

Segmentation Rules eXchange (SRX)

Segmentation Rules eXchange (SRX) is the vendor-neutral standard, originally published by LISA (Localization Industry Standards Association), for describing how translation and other language-processing tools segment text for processing.

SRX is an XML vocabulary that allows Translation Memory (TM) and other linguistic tools to describe the language-specific processes by which text is broken into segments (usually sentences or paragraphs) for further processing.

SRX uses regular expressions to indicate the conditions under which a given text needs to be broken or kept as is.

SRX version 2.0 was officially accepted as an OSCAR standard in April 2008.

SRX specification is currently available at http://www.ttt.org/oscarStandards/srx/.

Regular Expressions

Regular expressions, also known as *patterns*, are formulas that describe text strings using a specially designed syntax.

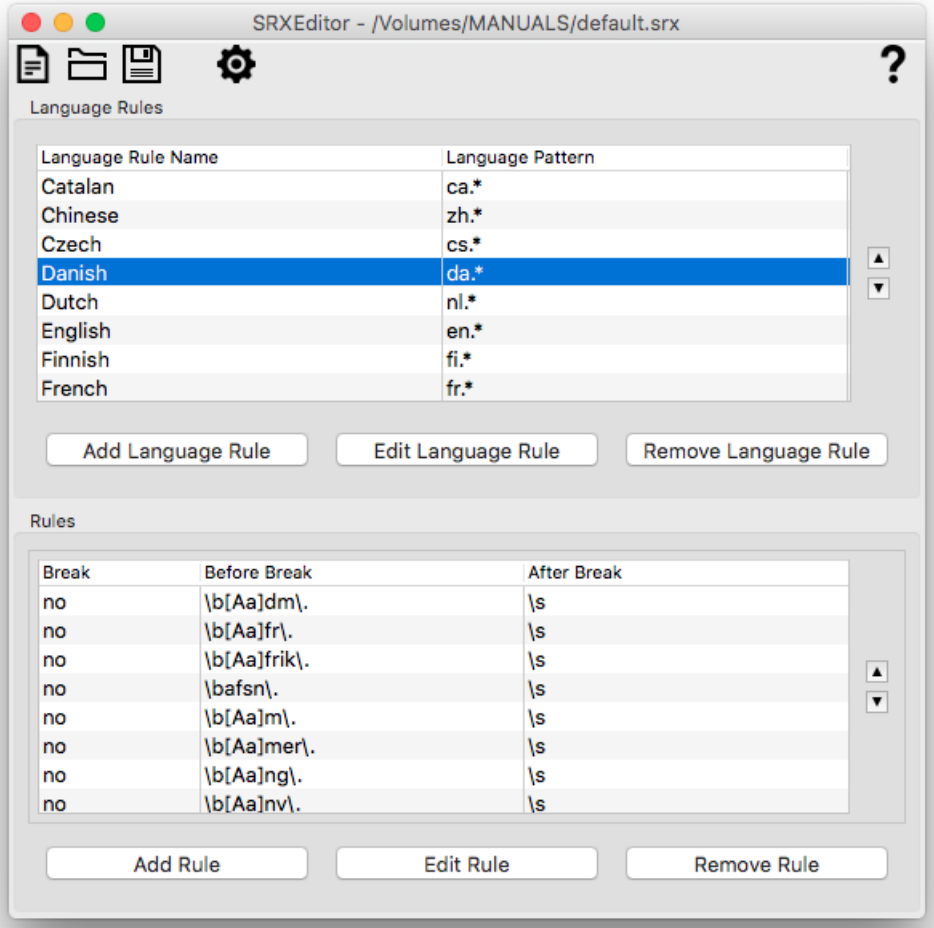
There are different kinds of regular expressions, usually associated with the programming language that uses them. The most common ones are Perl and Java regular expressions. Although they are very similar, small differences in syntax can lead to very different results when matching text against the same patterns.

SRX (Segmentation Rules eXchange) uses a subset of ICU (Internationalization Components for Unicode) regular expressions. The allowed subset is described in SRX 's specification document available for download at http://www.ttt.org/oscarStandards/srx/.

User Interface

SRXEditor GUI

The following picture portrays SRXEditor:



Menus

File Menu

	New File	Create a new SRX file.
	Open File	Open an existing SRX file for editing.
	Save File	Save current open SRX file to disk.
	Save File As...	Save current open SRX file to disk under a new name.
	Exit/Quit	Close the program.

Options Menu

	Language	Open a sub-menu for selecting the user interface language.

Help Menu

?	SRXEditor Help	Display SRXEditor User Guide in the default PDF viewer.
	About...	Display a dialog with copyright and program version information.

Common Tasks

Create SRX File

Create an SRX file that can be used for segmenting documents by translation tools that support SRX 2.0.

Procedure

1. In the **File** menu, select **New File** or click the button.
2. Click the **Add Language Rule** button to display the **Language Rule** dialog.
3. In the **Language Rule Name** text box, enter a meaningful name for the language rule.
4. In the **Language Pattern** text box, enter a regular expression to indicate the languages supported by the new rule.
5. Click the **Accept** button to add the language rule.

   The new language rule appears in the **Language Rules** table of the editor.

6. Follow these steps as many times as necessary to populate the language rule with segmentation rules:
   a. Click the **Add Rule** button to display the **Rule Editor** dialog.
   b. Select the **Breaks Segment** check box if the new segmentation rule breaks text. Leave it unselected otherwise.
   c. In the **Before Break** text box, enter a regular expression that describes the text that appears before the break point or break exception.
   d. In the **After Break** text box, enter a regular expression that describes the text that appears after the break point or break exception.
   e. Click the **Accept** button to add the new segmentation rule.

      The new rule appears in the **Rules** table of the editor.

7. In the **File** menu, select **Save File** or click the button to specify a file name for the SRX file and save it.

Results

A new SRX 2.0 file is generated and saved under the selected name.

Edit existing SRX File

Modify the content of an existing SRX file to segment text appropriately when using a tool that supports SRX 2.0

Procedure

1. In the **File** menu, select **Open File** or click the button.
2. Locate and open the SRX file to be modified.
3. Select a language rule for editing.
   - In the **Language Rules** table, click on an existing language rule.
   - Click the **Add Language Rule** button to add a new, empty, language rule.

The segmentation rules associated with the selected language rule are displayed in the **Rules** table.

4. Update the segmentation rules.
   - Click the **Add Rule** button to add a new segmentation rule.
   - In the **Rules** table, click on an existing segmentation rule to select it. Click the **Edit Rule** button to modify its content or click the **Remove Rule** button to delete the selected rule.
5. In the **File** menu, select **Save File** or click the button to save your changes. Use the **Save File As...** option to save the modified SRX file under a different name.

Results

The selected SRX file is saved with its content updated.

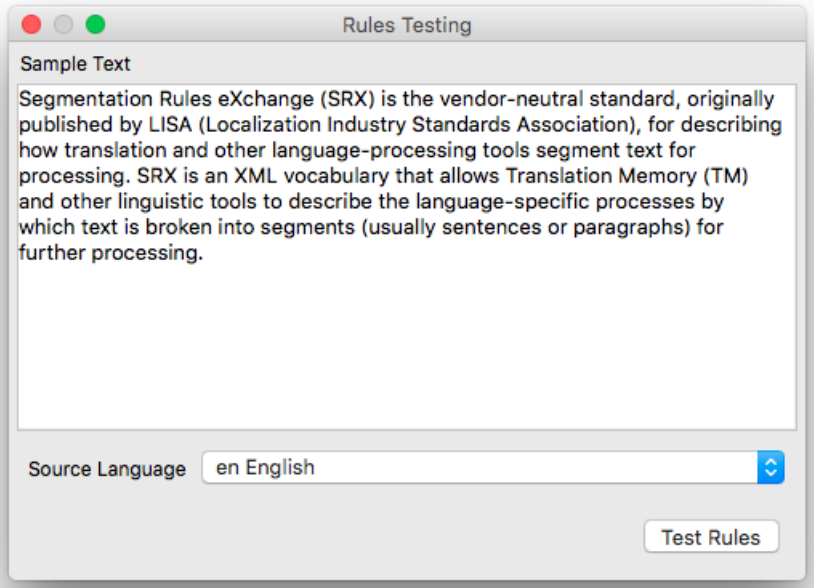
Test Segmentation Rules

Test whether the rules contained in an SRX file can be used to segment text as expected or not.

Procedure

1. Click the **Test Rules** button ( ) from the tool bar to display the **Rules Testing** dialog.

   The following dialog appears:



2. In the **Sample Text** box, type or paste text to segment.
3. Select the language of the sample text in the **Source Language** drop-down list.
4. Click the **Test Rules** button.

Results

An HTML table with the sample text segmented according to the specified rules set is displayed in a new window.

Glossary

Language Rule

A collection of segmentation rules associated with a set of languages defined using a regular expression.

Regular Expression

Formula or expression that describes text strings using a specially defined syntax.

Segmentation Rule

A pair of regular expressions that defines if a given text string has to be broken at segmentation time or if it is an exception that needs to be left untouched.

SRX

Segmentation Rules eXchange (SRX) is an XML-based open standard, published by LISA (Localization Industry Standards Association), for describing how translation and other language-processing tools segment text for processing.