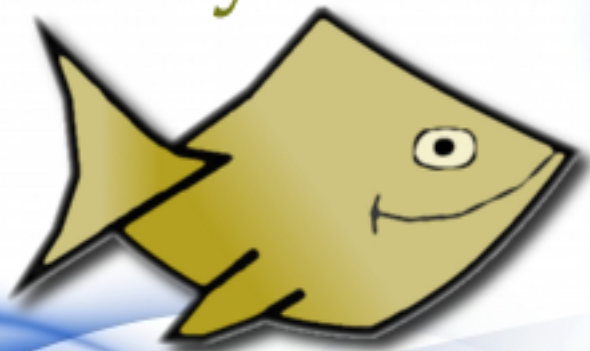


Anchovy User Guide

Anchovy



Copyright © 2009-2010 Maxprograms

Copyright © 2009 - 2010 Maxprograms

Contents

Introduction.....	5
Supported Platforms.....	5
Requirements.....	5
Supported Formats.....	5
Glossary Markup Language (GlossML).....	5
Comma Separated Values (CSV).....	5
TermBase eXchange (TBX).....	6
Translation Memory eXchange (TMX).....	6
User Interface.....	7
Anchovy GUI.....	7
Menus.....	8
File Menu.....	8
Edit Menu.....	8
Options Menu.....	8
Tasks Menu.....	8
Help Menu.....	9
Common Tasks.....	11
Create Glossary.....	11
Edit an Existing Glossary.....	11
Convert CSV File to GlossML Format.....	12
Extract Terms From a Document.....	13
Convert GlossML File to Other Formats.....	14
Export Glossary as HTML.....	15
Export Glossary as TMX.....	15
Export Glossary as TBX.....	15
Export Glossary as CSV.....	15
XSL Transformation.....	16
Glossary.....	17

Introduction

Anchovy is a cross-platform glossary editor, based on the open GlossML (Glossary Markup Language) format.

Anchovy simplifies glossary data exchange, as it is able to import CSV files and exports in CSV, HTML, TBX and TMX formats, the most common ones used in localization industry.

With Anchovy's built-in XSL Transformation engine, it is possible to export and import glossary data to and from almost any XML-based format.

Supported Platforms

- Microsoft Windows XP/Vista/7
- Mac OS X 10.4/10.5/10.6 (Tiger, Leopard and Snow Leopard)
- Linux with GTK2 (i386 and x86_64)

Requirements

- Java 1.5 or better from Sun Microsystems (available at <http://www.java.com>)
- Mozilla XULRunner is required to run on Linux systems.

Supported Formats

Glossary Markup Language (GlossML)

GlossML is an XML-based vocabulary specifically designed for containing glossaries that can be used for storing monolingual and multilingual lists of terms and, optionally, their definitions.

A distinctive aspect of GlossML vocabulary is its extreme simplicity. It only has 6 elements and 4 attributes. This is possible because it focuses solely on holding glossary data. It is not intended for terminology exchange.

The GlossML specification and related materials (XML Schema and examples) are licensed under the Creative Commons Attribution-No Derivative Works 3.0 Unported License. This means that anyone can use and distribute the GlossML format without paying royalties of any kind.

GlossML specification is available at <http://www.maxprograms.com/glossml/glossml.pdf>.

Comma Separated Values (CSV)

A comma-separated values (CSV) file is used for the digital storage of data structured in a table of lists form, where each associated item (member) in a group is in association with others also separated by the commas of its set. Each line in the CSV file corresponds to a row in the table. Within a line, fields are separated by commas, each field belonging to one table column.

Since it is a common and simple file format, CSV files are often used for moving tabular data between two different computer programs, for example between a database program and a spreadsheet program.

Very often, characters other than commas are used to separate columns. Also, it's also often necessary to enclose column data in special delimiters when they contain the characters used as column separators.

TermBase eXchange (TBX)

Term Base eXchange (TBX) is the open, XML-based standard for exchanging structured terminological data that has been approved as an international standard by LISA and ISO. It is also known as ISO 30042 standard.

TBX is a framework that allows custom definition of terminology databases. It requires two files per terminology database: one file that defines the constraints and data categories to be used (the XCS file) and a second file containing the actual data (the TBX file).

TBX is available in two flavors:

- | | |
|------------------|--|
| TBX | The official version released by ISO and LISA together. This version is based on two DTDs with a large set of XML elements and attributes. An extremely rich set of data categories is included in the default XCS file, designed to support multiple subjects through a large number of available fields. |
| TBX Basic | A limited subset of TBX, which intends to simplify the complex set of options included in the standard one. This separate version published only by LISA has its own DTD and official data categories defined in a separate XCS file. |

TBX specification is available at http://www.lisa.org/fileadmin/standards/TBX_2008_10_29.pdf

Translation Memory eXchange (TMX)

TMX (Translation Memory eXchange) is the vendor-neutral open XML standard for the exchange of Translation Memory (TM) data created by Computer Aided Translation (CAT) and localization tools.

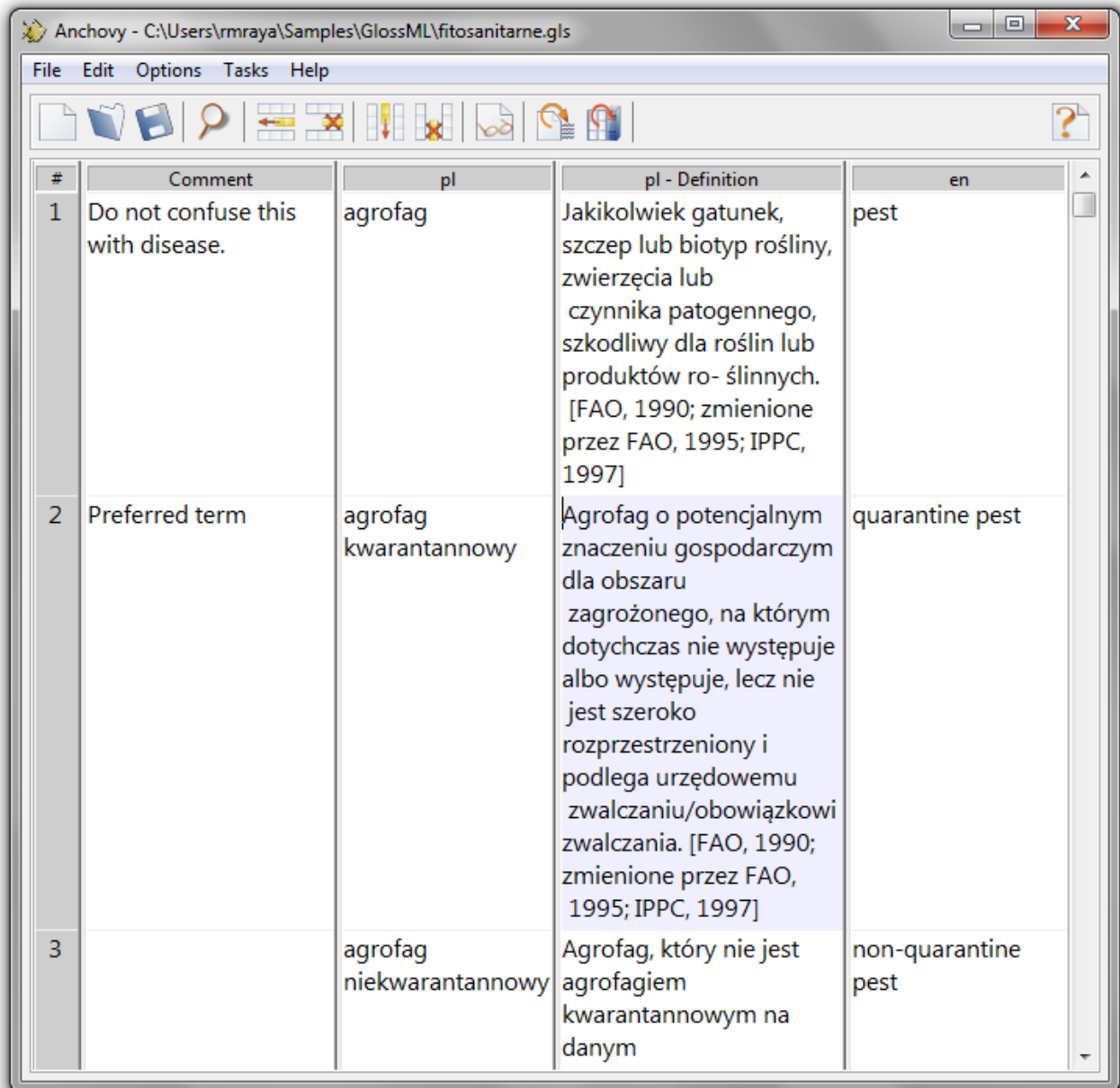
The purpose of TMX is to allow easier exchange of translation memory data between tools and/or translation vendors with little or no loss of critical data during the process.

TMX specification is available at <http://www.lisa.org/tmx/tmx.htm>.

User Interface

Anchovy GUI

The following picture portrays Anchovy:






The screenshot shows the Anchovy GUI window with the following table content:






#	Comment	pl	pl - Definition	en
1	Do not confuse this with disease.	agrofag	Jakiegolwiek gatunek, szczep lub biotyp rośliny, zwierzęcia lub czynnika patogennego, szkodliwy dla roślin lub produktów roślinnych. [FAO, 1990; zmienione przez FAO, 1995; IPPC, 1997]	pest
2	Preferred term	agrofag kwarantannowy	Agrofag o potencjalnym znaczeniu gospodarczym dla obszaru zagrożonego, na którym dotychczas nie występuje albo występuje, lecz nie jest szeroko rozprzestrzeniony i podlega urzędowemu zwalczaniu/obowiązkowi zwalczania. [FAO, 1990; zmienione przez FAO, 1995; IPPC, 1997]	quarantine pest
3		agrofag niekwarantannowy	Agrofag, który nie jest agrofagiem kwarantannowym na danym	non-quarantine pest

Menus

File Menu

	New File	Create a new GlossML file.
	Open File	Open an existing GlossML file for editing.
	Save File	Save current open GlossML file to disk.
	Save File As...	Save current open GlossML file to disk under a new name.
	Export as HTML	Save the content of current open GlossML file in HTML format.
	Export as TMX	Save the content of current open GlossML file in TMX format.
	Export as TBX	Save the content of current open GlossML file in TBX format.
	Export as CSV	Save the content of current open GlossML file in CSV format.
	Exit/Quit	Close the program.




Edit Menu

	Search/Replace	Open a dialog for searching/replacing text.
	Add Row	Add a new row to current open GlossML file.
	Delete Row	Delete selected row from current open GlossML file.
	Add Column	Add a new column to current open GlossML file.
	Delete Column	Delete selected column from current open GlossML file.


Options Menu

	Language	Display a submenu for selecting the language to use in the user interface.
	Font Settings	Display a dialog for selecting the font to use in the application.

Tasks Menu

	View as HTML	Open current GlossML file in a web browser.
	Term Extraction	Open a dialog for extracting common terms from a document.
	Convert CSV File to GlossML Format	Open a dialog for converting CSV and other delimited files to GlossML format.
	XSL Transformation	Display a dialog for applying XSL transformations to an XML file.

Help Menu

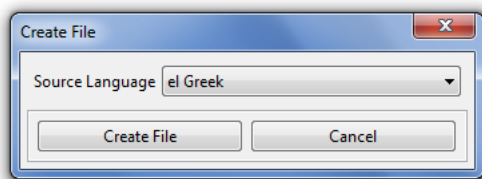
	Anchovy Help	Display Anchovy User Guide in the default PDF viewer.
	About...	Display a dialog with program version information.


Common Tasks

Create Glossary

Create a new glossary in GlossML format.

1. In the **File** menu, select **New File** or click the  button.
The following dialog appears:








2. In the **Create File** dialog, select the source language for the new glossary using the **Source Language** drop-down list.
3. Click the **Create File** button to create an empty GlossML file and open it in Anchovy.
4. Add content to the glossary.
5. In the **File** menu, select **Save File** or click the  button to save the newly created GlossML file.


A new GlossML file with columns for terms and definitions for the specified source language is created and displayed in Anchovy.


Edit an Existing Glossary

Modify the content of an existing glossary in GlossML format.

1. In the **File** menu, select **Open File** or click the  button.
2. Locate and open the GlossML file to be modified.
3. Select a cell and modify its content or use one of the options in the [Edit](#) menu or their equivalent buttons in the tool bar to perform additional tasks.

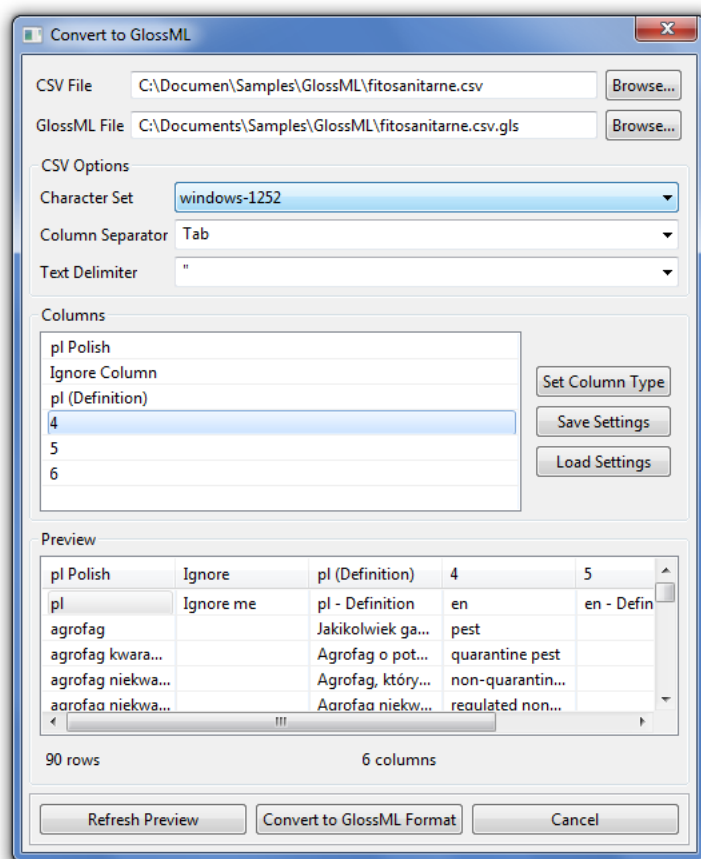
Options	Description
 Search/Replace	Use this option to search for a cell containing the specified text and optionally replace that text with a different version.
 Add Row	Use this option to add a new blank row to the glossary in which to insert a new glossary term.
 Delete Row	Use this option to delete a row from a glossary, removing a term, its definition and possible translations.
 Add Column	Use this option to add a new column to the glossary. Available column types are: Comment Column A column in which comments about a term are added. This column, when available, is the first one on the left side.

Options	Description
Term Column	A column for inserting terms in the language selected when the column is added to the glossary.
Definition Column	A column for defining the associated term in the language selected when the column is added to the glossary.
 Delete Column	Delete the selected column from the glossary.

- 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 GlossML file under a different name.

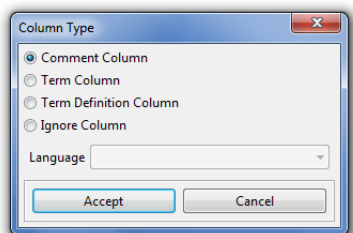
Convert CSV File to GlossML Format

- In the **Tasks** menu, select **Convert CSV File to GlossML Format** or click the  button. The following dialog appears:



- Type the name of the CSV file in the **CSV File** text box or use the **Browse...** button next to it for selecting an existing CSV file from the file system. If you enter the name of the file manually, click the **Refresh Preview** button to load its content in the **Preview** panel.
- Type the name of the GlossML file to create in the **GlossML File** text box or use the **Browse...** button next to it to select a file name and location.


- Adjust the values of **Character Set**, **Column Separator** and **Text Delimiter** drop down lists in the **CSV Options** panel until the data is properly displayed in the **Preview** panel. If necessary, use the **Refresh Preview** button to update the display after changing a value.
- Select a column from the **Columns** list and click on the **Set Column Type** button.
The following dialog appears:

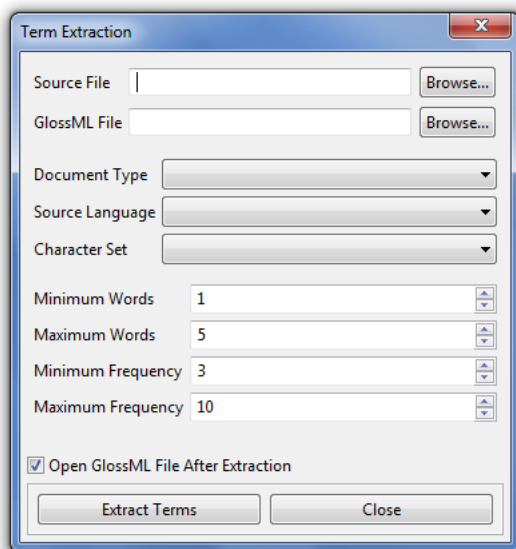


- In the **Column Type** dialog, select the type of column and, if required, select the language of the column.
- Click **Accept** button to close the **Column Type** dialog.
The corresponding column type and language will be displayed in the **Columns** list and the **Preview** panel will be updated.
- Repeat the previous 3 steps until all columns have a type assigned.
- Click the **Convert to GlossML Format** button.

A new GlossML file is created and stored in the selected location.

Extract Terms From a Document

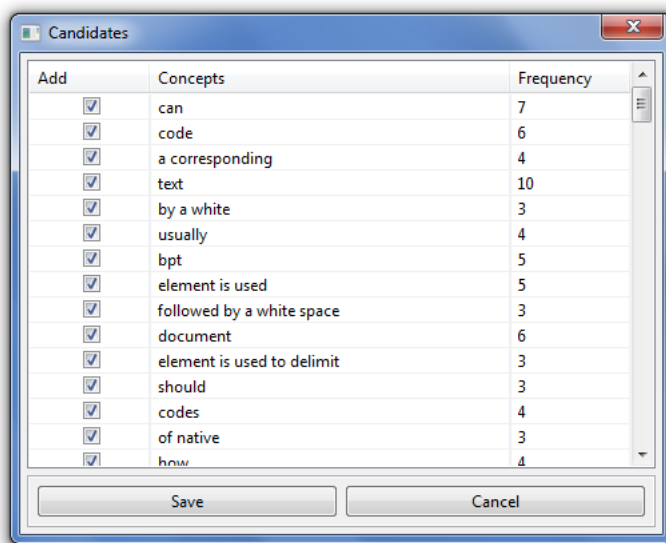
- In the **Tasks** menu, select **Term Extraction** or click the  button.
The following dialog appears:



- Type the name of the document to be analyzed in the **Source File** or use the **Browse...** button next to it to select a document from the file system.

3. Type the name of the GlossML file in the **GlossML File** text box or use the **Browse...** button next to it to select a file name and location.
4. Select the document type in the **Document Type** drop-down list or correct the type automatically selected by the program if necessary.
5. Select the language of the source document in the **Source Language** drop-down list.
6. Select the character set of the document in the **Character Set** drop-down list or correct the one automatically selected by the program if necessary.
7. Indicate the minimum number of words a term must contain in the **Minimum Words** selector.
8. Indicate the maximum number of words a term can contain in the **Maximum Words** selector.
9. Indicate in the **Minimum Frequency** selector the minimum number of times a candidate term must appear in the source document.
10. Indicate in the **Maximum Frequency** selector the maximum number of times a candidate term can appear in the source document.
11. Check the **Open Glossary After Extraction** box if you want to edit the generated GlossML file in Anchovy after finishing the extraction process.
12. Click the **Extract Terms** button.

The source document is analyzed and a list of candidate terms is displayed in the following dialog:



13. Review the list of candidate terms and uncheck the box that appears in the leftmost column to discard the candidates that you want to ignore. The candidates list can be sorted by status, alphabetically or by frequency clicking on the corresponding column labels at the top of the list.
14. Click on the **Save** button to store the selected terms in the previously indicated GlossML file.

A GlossML file containing the selected terms is generated and, optionally, opened in Anchovy.


Convert GlossML File to Other Formats

Anchovy facilitates data exchange by allowing easy conversion of GlossML files to other formats.

Conversion of GlossML files to TMX, TBX and HTML formats is done by applying XSL transformations. The XSL stylesheets used in the conversion are stored in /stylesheets folder of the installation directory and can be customized by the end user.


Conversion to CSV is also flexible, allowing the selection of column separators, text delimiters and column filtering at export time.

Export Glossary as HTML

1. In the **File** menu, select **Open File** or click the  button.
2. Locate and open the GlossML file to be exported.
3. In the **File** menu, select **Export as HTML**.
4. Select a file name and location for storing the generated HTML file.


An HTML file is generated and saved in the selected location.

Export Glossary as TMX

1. In the **File** menu, select **Open File** or click the  button.
2. Locate and open the GlossML file to be exported.
3. In the **File** menu, select **Export as TMX**.
4. Select a file name and location for storing the generated TMX file.


A TMX file is generated and saved in the selected location.

Export Glossary as TBX

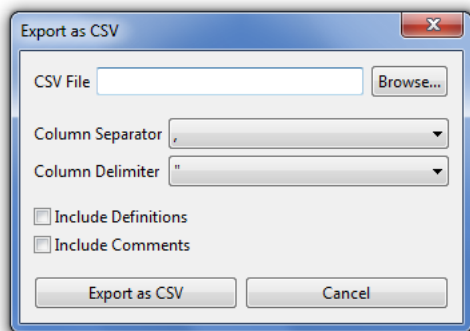
1. In the **File** menu, select **Open File** or click the  button.
2. Locate and open the GlossML file to be exported.
3. In the **File** menu, select **Export as TBX**.
4. Select a file name and location for storing the generated TBX file.

A TBX file is generated and saved in the selected location.

Export Glossary as CSV

1. In the **File** menu, select **Open File** or click the  button.
2. Locate and open the GlossML file to be exported.
3. In the **File** menu, select **Export as CSV**.

The following dialog appears:



4. Type the name of the CSV file in the **CSV File** text box or use the **Browse...** button next to it to select a file name and location.

5. Select the column separator to use in the CSV file from the **Column Separator** drop-down list.
6. Select the column delimiter to use in the CSV file from the **Column Delimiter** drop-down list.
7. Check the **Include Definitions** box if you want to include term definitions in the exported CSV file.
8. Check the **Include Comments** box if you want to include a column with comments in the exported CSV file.
9. Click the **Export as CSV** button.

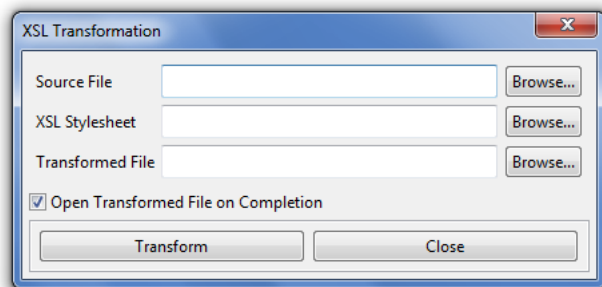
A CSV file with selected delimiters and data is written in the selected location using the UTF-16 character set.

XSL Transformation

Transform an XML document to a different format by applying an XSL transformation.

1. In the **Tasks** menu, select **XSL Transformation**.

The following dialog appears:



2. Type the name of the document to be transformed in the **Source File** text box or use the **Browse...** button next to it for selecting an existing document from the file system.
3. Type the name of the XSL Stylesheet to apply in the **XSL Stylesheet** text box or use the **Browse...** button next to it for selecting the stylesheet from the file system.
4. Type the name of the document to generate with the transformation process in the **Transformed File** text box or use the **Browse...** button next to it to specify the name and location of the resulting file.
5. Check the **Open Transformed File on Completion** box if you want to open the transformed file in the default viewer after the transformation is finished. Leave this check box empty otherwise.
6. Click the **Transform** button.

The selected XSL stylesheet is applied to the source XML document and the result of the transformation is stored in the indicated file. Transformation results are displayed in the default viewer configured in the operating system if the corresponding check box is selected.

Glossary

Definitions of terms used in this book.

Character Set	A character set (sometimes referred to as code page) is a collection of characters that are associated with a sequence of natural numbers in order to facilitate the storage of text in computers and the transmission of text through telecommunication networks.
Column Delimiter	Character used to delimit the content of a data column in CSV files. Double quotes are frequently used as delimiters. Delimiters are required whenever the column text contains the character used as Column Separator.
Column Separator	Character used to separate data columns in a CSV file. Usually a comma.
CSV	CSV (Comma Separated Values) is a standard file format used to store tabular data.
GlossML	Glossary Markup Language (GlossML) is an XML vocabulary specifically designed for containing glossaries used in translation/localization industry.
TBX	TBX (TermBase eXchange) is the open, XML-based standard for exchanging structured terminological data. First released by LISA in May, 2002, TBX was submitted to the International Organization for Standardization (ISO) on February 21, 2007, for adoption as an ISO standard.
TML	Terminological Markup Language (TML) is an XML framework for describing a terminological data collection. In the case of a TBX TML, the framework consists of two files: a DTD that specifies the elements and attributes to use in the XML container for terminology data and an XCS file for describing the data categories and constraints.
TMX	Translation Memory eXchange (TMX) is an open standard published by LISA (Localization Industry Standards Association). The purpose of TMX is to allow easier exchange of translation memory data between tools and/or translation vendors with little or no loss of critical data during the process.
XCS	XCS (eXtensible Constraint Specification) is an XML vocabulary that defines the data categories and their constraints for a specific TBX TML.

