SDL MultiTerm 2011 Extract
Because Brand Matters.
COPYRIGHT

SDL MULTITERM EXTRACT TOOLS
USER GUIDE

Copyright • 2009-2011 SDL plc.

All rights reserved. Unless explicitly stated otherwise, all intellectual property rights including those in copyright in the content of this website and documentation are owned by and controlled for these purposes by SDL PLC. Except as otherwise expressly permitted hereunder or in accordance with copyright legislation, the content of this site, and/or the documentation may not be copied, reproduced, republished, downloaded, posted, broadcast or transmitted in any way without the express written permission of SDL PLC.

This product may include open source or similar software designated: Hunspell distributed under GNU Lesser General Public License version 3; Sharpziplib and Spring.net distributed under GNU General Public License version 3; ICU distributed under IBM ICU License; Log4Net, Xalan and Xerces distributed under Apache License version 2.0; Wix distributed under Common Public License version 1.0; and SQLite which is public domain and requires no license for this distribution.

Trados, MultiTerm, SDLX and TranslationZone are registered trademarks of SDL plc. Translator’s Workbench, Trados Studio and TagEditor are trademarks of SDL plc. All other trademarks are the property of their respective owners. The names of other companies and products mentioned herein may be the trademarks of their respective owners. Unless stated to the contrary, no association with any other company or product is intended or should be inferred.

Although SDL takes all reasonable measures to provide accurate and comprehensive information about the product, this documentation is provided as-is and all warranties, conditions or other terms concerning the documentation whether express or implied by statute, common law or otherwise (including those relating to satisfactory quality and fitness for purposes) are excluded to the extent permitted by law.

Information in this documentation, including any URL and other Internet Web site references, is subject to change without notice. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of SDL plc.

This guide ships with SDL MultiTerm Extract 2011.

September 2011
TABLE OF CONTENTS

GENERAL INFORMATION 1

Chapter 1  SDL MultiTerm Extraction Tools 1

  About The SDL Extraction Tools 2
  Statistical Extraction using SDL MultiTerm Extract 2

SDL MULTITERM EXTRACT 1

Chapter 2  About SDL MultiTerm Extract 1

  About Statistical Extraction in SDL MultiTerm Extract 2
  Supported Languages 2
  Supported File Formats 2
    Export Formats 4
  Related Documentation 5
    Accessing Documentation 5

Chapter 3  Working with SDL MultiTerm Extract 1

  SDL MultiTerm Extract Projects 2
    Project Types 2
    Overview of the Term Extraction Process 3
Overview of the Term Extraction Process for QA Projects 5

Creating a Project 6
   Before You Begin 6

Saving a Project 21

Importing Terms into Translation Projects 22

Extracting Terms 23
   Automatically Extracting Terms 23
   Manually Extracting Terms 24
   Text Window 25

SDL MultiTerm Extract Interface 26
   Term Window 27
   Term Properties Window 29

Validating Extracted Terms 35

Chapter 4  Exporting Terms in SDL MultiTerm Extract 1

About Exporting Extracted Terms 2
   Export Definitions 2
   Before You Begin 2

Exporting Terms 3

Exporting Using Export Definitions 11
   Exporting Using an Existing Export Definition 11
   Exporting Using a New Export Definition 12

Exporting Suspect Sentences in a QA Project 15

Using Filters when Exporting 16

Chapter 5  SDL MultiTerm Extract Reference 1

Settings 2
   Preferences 2
   Project Settings 3
   File Format Settings 9
QA Statistics 11

Keyboard Shortcuts 13
This chapter provides an overview of the SDL MultiTerm Extraction Tools. Sections include:

- About The SDL Extraction Tools
- Statistical Extraction using SDL MultiTerm Extract
ABOUT THE SDL EXTRACTION TOOLS

Welcome to the SDL MultiTerm Extract Tools Guide. This guide explains the method of statistical terminology extraction using SDL MultiTerm Extract.

This allows you to extract candidate terms to add to a new or existing termbase. It reduces the time and effort involved in generating a termbase by automating the process where terms are identified and entered into a central repository for review and approval. This is especially useful when creating a termbase from large volumes of documents or from new information that is constantly being updated.

STATISTICAL EXTRACTION USING SDL MULTITERM EXTRACT

SDL MultiTerm Extract uses a statistical extraction method to determine the frequency of the appearance of candidate terms. It extracts term candidates and, for multilingual termbases, their probable translations found in sentences, incomplete sentence fragments and strings of code. The extracted terms are presented as a term candidate word or phrase.

After running the extraction process, SDL MultiTerm Extract displays a term candidate list. Following review and validation, term candidates become term lists. In the SDL MultiTerm Extract interface, you can quickly and easily validate terms. You can then use the multiple export options to incorporate extracted terms into your existing MultiTerm terminology management system. Terms can be exported to a MultiTerm termbase, MultiTerm XML format or a tab-delimited format. This allows, for example, translators to access those terms when using the Translator’s Workbench termbases.

SDL MultiTerm Extract supports all languages, including those of the Far East. If you are using a non-Western languages, you must have support for the relevant language installed on your computer.
ABOUT SDL MULTI TER M EXTRACT

This chapter provides an overview of the statistical extraction method using SDL MultiTerm Extract. Sections include:

- About Statistical Extraction in SDL MultiTerm Extract
- Supported Languages
- Supported File Formats
- Related Documentation
ABOUT STATISTICAL EXTRACTION IN SDL MULTITERM EXTRACT

SDL MultiTerm Extract uses statistical extraction to extract term candidates and, for multilingual termbases, their probable translations and present them in a term candidate list. You can extract terms from monolingual or bilingual documents and from translation memories and then quickly and easily validate the terms.

The various export options enable you to incorporate extracted terms into existing SDL MultiTerm Extract termbases or to export terms to MultiTerm XML format or to tab-delimited format. You can also check for terminology consistency between the termbase and a translated file.

SDL MultiTerm Extract allows you to do the following:

- Create term candidate lists from monolingual documents.
- Create bilingual term candidate lists from bilingual file formats.
- Update existing multilingual termbases by creating new translations for terms already stored in MultiTerm.
- Analyze the quality of termbases and documents by comparing the terms found in documents with those stored in the termbase.
- Compile bilingual dictionaries from bilingual file formats.
- Manually extract terms from the source document(s) to complement and enhance the automatic term extraction process.
- Export extracted terms to MultiTerm, MultiTerm XML or a tab-delimited format.

SUPPORTED LANGUAGES

SDL MultiTerm Extract is fully Unicode-compliant and supports all Unicode-compliant languages. However, to render non-Western languages successfully in SDL MultiTerm Extract, you must have support for the relevant language installed on your computer.

SUPPORTED FILE FORMATS

SDL MultiTerm Extract supports term extraction from a large range of file formats. The supported file formats depend on the type of project that you are using.
The file formats supported by the monolingual term extraction can be used in a bilingual project, but no translations are extracted. The formats can be used when you want to find source terms but you must add your own translation for any terms that are extracted.

<table>
<thead>
<tr>
<th>Project</th>
<th>Supported File Types</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monolingual Term Extraction</strong></td>
<td>TXT – Plain text&lt;br&gt;RTF – Rich text format&lt;br&gt;Microsoft Office Word (<em>.doc)&lt;br&gt;HTML, HTM, JSP, ASP, ASPX&lt;br&gt;SGML, SGM, XML&lt;br&gt;Ventura (</em>.txt)&lt;br&gt;Adobe PageMaker (<em>.txt)&lt;br&gt;QuarkXPress (</em>.qsc, *.txg, <em>.ttg, <em>.tag)&lt;br&gt;TTX – TRADOSTag file format&lt;br&gt;Adobe InDesign (</em>.isc)&lt;br&gt;Microsoft Office PowerPoint (</em>.ppt, *.pps, <em>.pot)&lt;br&gt;Microsoft Office Excel (</em>.xls, *.xlt)</td>
</tr>
<tr>
<td><strong>Bilingual Term Extraction</strong></td>
<td>BILINGUAL FORMATS:&lt;br&gt;TMX – Translation memory exchange format&lt;br&gt;TMW – Translator’s Workbench&lt;br&gt;Translator’s Workbench TXT format (<em>.txt)&lt;br&gt;TTX – TRADOSTag files&lt;br&gt;MONOLINGUAL FORMATS:&lt;br&gt;TXT – Plain text&lt;br&gt;RTF – Rich text format&lt;br&gt;Microsoft Office Word (</em>.doc)&lt;br&gt;HTML, HTM, JSP, ASP, ASPX&lt;br&gt;SGML, SGM, XML&lt;br&gt;Ventura (<em>.txt)&lt;br&gt;Adobe PageMaker (</em>.txt)&lt;br&gt;QuarkXPress (*.qsc, <em>.txg, <em>.ttg, <em>.tag)&lt;br&gt;Translator’s Workbench RTF format (</em>.rtf)&lt;br&gt;Adobe InDesign (</em>.isc)&lt;br&gt;Microsoft Office PowerPoint (</em>.ppt, *.pps, <em>.pot)&lt;br&gt;Microsoft Office Excel (</em>.xls, *.xlt)</td>
</tr>
<tr>
<td>Supported File Formats</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Export Formats</strong></td>
<td></td>
</tr>
</tbody>
</table>

You can export extracted terms directly to a MultiTerm termbase, a MultiTerm XML export file or a tab-delimited text file.
RELATED DOCUMENTATION

The following documentation may be useful when working with SDL MultiTerm Extract:

- SDL MultiTerm Extract online help. There is also a SDL MultiTerm Extract tutorial installed with the product and is accessible from the Start menu.
- Trados Studio online help.

Accessing Documentation

- All guides are available in PDF format in the docs subfolder of your product DVD, and in the Doc subfolder of your installation.

SDL MultiTerm Extract is available for users of SDL MultiTerm Desktop. Contact your SDL sales representative for details.
This chapter provides information about using SDL MultiTerm Extract. Sections include:

- SDL MultiTerm Extract Projects
- Extracting Terms
- Importing Terms into Translation Projects
- SDL MultiTerm Extract Interface
- Validating Extracted Terms
SDL MULTITERM EXTRACT PROJECTS

SDL MultiTerm Extract projects allow you to extract terms from existing files. Those terms can then be reviewed, validated and exported to a termbase or other available formats to assist in translation.

Project Types

There are five types of project in SDL MultiTerm Extract:

- **Monolingual Term Extraction Project**
- **Bilingual Term Extraction Project**
- **Dictionary Compilation Project**
- **Translation Project**
- **QA Project**

Use the **New Project Wizard** in SDL MultiTerm Extract to create any of the projects.

**NOTE**

SDL MultiTerm Extract support a wide range of file formats. These formats vary depending on which project is created. See “Supported File Formats” on page 2-2.

<table>
<thead>
<tr>
<th>Project</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monolingual Term Extraction Project</strong></td>
<td>Extracts terminology from a text in one language.</td>
</tr>
<tr>
<td><strong>Bilingual Term Extraction Project</strong></td>
<td>Extracts terminology from monolingual formats. Extracts and translates from bilingual formats.</td>
</tr>
<tr>
<td><strong>Dictionary Compilation Project</strong></td>
<td>Extracts a dictionary from an existing translation memory file or a bilingual file which has been translated using Translator’s Workbench.</td>
</tr>
<tr>
<td><strong>QA Project</strong></td>
<td>Analyzes and allows you to improve the quality of termbases and documents.</td>
</tr>
<tr>
<td><strong>Translation Project</strong></td>
<td>Provides suggestions from existing texts and allows you to enter translations to a termbase.</td>
</tr>
</tbody>
</table>
About Translation Projects

A Translation Project updates an existing termbase in MultiTerm by creating new translations in one or more languages for corresponding terms already stored in the termbase. You will need a text in a supported file format and the MultiTerm termbase.

For example, if your termbase contains English and Spanish indexes, and already has the English term “Door”, and your translated file contains the equivalent term in Spanish, the Translation Project can update the termbase automatically with this term.

About QA Projects

A QA Project analyzes and improves the quality of termbases and documents by comparing the terms found in documents with those stored in the termbase.

A QA Project allows you to:

- Monitor the usage of terms found in project files and the project termbase and their respective translations. This allows you to check that the terms used in the document are those that are contained in the termbase.

- Evaluate the translation quality of documents on the basis of the stored terminology.

Each QA Project provides a QA report. A QA report provides statistical information on the terms in project files and existing terms from the MultiTerm termbase in the project.

Overview of the Term Extraction Process

All SDL MultiTerm Extract projects are created using the New Project Wizard. This procedure varies depending on what type of project you are creating. Displayed below is an overview of creating a Monolingual Term Extraction Project, Bilingual Term Extraction Project, Dictionary Compilation Project or Translation Project.

Each step is explained in more detail later in this guide.

1. Create a new SDL MultiTerm Extract project by selecting File > New from the menu bar. The New Project Wizard allows you to select the project type, add files, customize settings and specify stopword lists. See “Creating a Project” on page 3-6.

2. Start the term extraction process by selecting Project > Process from the menu bar. See “Extracting Terms” on page 3-23.
3 View the extracted term list in the **Term** window. For **Translation Projects**, the translations for the source terms may be located in either in the input files or in the MultiTerm termbase. See “Term Window” on page 3-27.

4 In the **Term properties** window you can check each word, and for **Translation Projects**, you can select a term and add translations.

5 Validate terms using the **Concordance** window for reference. See “Validating Extracted Terms” on page 3-35.

6 Export validated terms by selecting **File > Export** the menu bar and using the **Export Wizard**. See “Exporting Terms in SDL MultiTerm Extract” on page 4-1.
Overview of the Term Extraction Process for QA Projects

To check documents and termbases using a QA Project:

2. Extract the terms. See “Extracting Terms” on page 3-23.
3. Check the QA statistics. See “QA Statistics” on page 5-11.
7. Process the project and check the QA statistics again.
CREATING A PROJECT

All SDL MultiTerm Extract projects are created using the New Project Wizard.

The example shows you how to create all of the different project types in SDL MultiTerm Extract. The procedure for creating a Monolingual Term Extraction Project, Bilingual Term Extraction Project and Dictionary Compilation Project is similar.

NOTE

If you are extracting terms from HTML, SGML, XML, ASP or JSP formats, See “File Format Settings” on page 5-9.

Before You Begin

If you intend to use a remote termbase in your project, the following information is required to define termbase connection settings in the New Project Wizard:

- MultiTerm Server computer details – determine the MultiTerm Server computer name that provides access to the termbases with which you want to work.
- MultiTerm Server connection type – find out whether your computer and the MultiTerm Server computer are linked by a standard LAN or a MultiTerm Anywhere connection. A standard LAN connection applies when both computers belong to the same physical domain. A MultiTerm Anywhere connection applies when the two computers belong to separate physical domains.
- MultiTerm Server login details – determine the user name and password for your MultiTerm Server login.

Contact your MultiTerm administrator for these details and for information about your termbase user rights. You must have sufficient rights to work with a remote termbase from SDL MultiTerm Extract.

If you intend to use a local termbase in your project, know the name and location before you open the New Project Wizard. You have exclusive access and unlimited user rights in all local termbases, and no login details are required.
NOTE

When you connect to a MultiTerm Server, the license restrictions as defined on the server are applied, so you might not be able to open the termbase because no free viewing seats are available on the server. In this case, wait until an appropriate seat becomes available and inform your MultiTerm administrator.

To create a new SDL MultiTerm Extract project:

1. Select File > New Project from the menu bar. The New Project Wizard opens and the Choose project type, name and location page is displayed. Use this page to define the project type, name and location.

2. Select your project type:
   - **Monolingual Term Extraction Project** – contains extracted terms from one or more single language texts.
   - **Bilingual Term Extraction Project** – contains extracted source and target terms from bilingual files (TTX, TMW, TMX). You can also use WinAlign export files and add monolingual files to bilingual term extraction projects.
   - **Translation Project** – used to update an existing MultiTerm termbase by offering new translations for terms already stored in the termbase and adding new term pairs.
   - **QA Project** – helps you monitor the use of terms in a document and a termbase and the way they are translated. It allows you to evaluate the translation quality of documents and termbases.
   - **Dictionary Compilation Project** – contains an alphabetically-sorted dictionary compiled from bilingual files (TTX, TMW, TMX). It includes source language terms as well as their target language translations.
3 Complete the following information:

- Enter a name for the project in the **Project Name** box.
- Click **Browse** to select a location for the SDL MultiTerm Extract project. By default SDL MultiTerm Extract places all projects in a *Projects* folder within the SDL MultiTerm Extract installation folder.

4 Click **Next**. The **Choose termbase and languages** page is displayed. This is where you select the project languages and optionally, a project termbase.

![New Project Wizard]

Selecting a project termbase allows you to perform the following actions:

- You can exclude the terms in the termbase, from the extraction process. This means that known terms are ignored by SDL MultiTerm Extract.
- Once the project is created and the terms have been extracted and validated, you can export the terms from your project into the project termbase. See “Exporting Terms in SDL MultiTerm Extract” on page 4-1.

5 If you do not want to add a termbase to your project, select **No Termbase** and then complete the **Source Language** and **Target Language** boxes by selecting the project source and target languages. Only the languages supported by your operating system are available for selection. The **Target Language** box is not displayed if you are creating a monolingual project. Now go to step 8.

**NOTE**

All Translation Projects and QA Projects must have a termbase in the project. For this reason, the **No Termbase** option for these projects is not available.
6 If you want to add a termbase to your project, select **Termbase** and click **Choose**. The **Choose Termbase** dialog box is displayed. A list of the termbases you have used previously is displayed in the **Termbases** box.

- If the termbase you want to use is listed here, select the termbase and click **OK**. If the termbase is server-based, the **Connect to MultiTerm Server** dialog box is displayed. Enter your login name and password for the server and click **OK**. Go to step 7.

- If you want to use a local termbase that is not in the list, click **Browse termbase**. The **Open Termbase** dialog box is displayed. Navigate to the required termbase, select the termbase and click **Open**. The termbase is added to the **Termbases** box. Select the termbase and click **OK**. Go to step 7.

- If you want to use a server-based termbase that is not in the list, click **Remote login**. The **Connect to MultiTerm Server** dialog box is displayed.

- Enter your user name and password for the server and click **More** to display additional server settings.

**NOTE**

MultiTerm server login details are available from your MultiTerm administrator. If you intend to export terminology directly to the remote termbase, make sure that the login you receive provides you with sufficient termbase access rights.
Enter the name of the MultiTerm Server computer.

Select Standard LAN or MultiTerm Anywhere, depending on the type of connection that exists between your computer and the MultiTerm Server computer. If both computers belong to the same physical domain, choose Standard LAN. This is the connection that is generally used. If both computers belong to separate physical domains, choose MultiTerm Anywhere.

If you selected MultiTerm Anywhere and your organization uses a proxy server, select Use proxy server within your LAN and complete the Address and Port boxes.

Click OK to save the server connection details and close the Connect to MultiTerm Server dialog box. If the login details you provided were correct, the termbases available on the server are added to the list in the Choose Termbase dialog box.

Select the required termbase from the list and click OK to close the Choose Termbase dialog box.

NOTE
You only need to define the remote termbase connection settings for your project once. Thereafter, each time you open the project, the Connect to MultiTerm Server dialog box is automatically displayed enabling you to log in and access the relevant termbase in a more direct fashion.

Under Termbase, complete the Source Language and Target Language boxes by selecting the project source and target languages. Only the languages supported by your operating system are available for selection. The Target Language box is not displayed if you are creating a monolingual project.
8 Click Next. The Add files to this project page is displayed. This is where you select the files to be searched for potential terminology.

- Click Browse to open the Add dialog box. Browse to location of the files you want to add to the project. Select a file or a group of files and click OK. The selected file or files are added to the Project Files list. Continue adding files until the list is complete.

- If you need to specify a settings file (*.ini) for the files you have selected, click File Format Settings. See “File Format Settings” on page 5-9.

- If you want to create the project using the default settings, click Finish.

- If you want to change the default settings before creating the project, go to step 9.
Click **Next**. The **Excluded terms** page is displayed. Use this page to define the exclusion and learning settings for your project. This page is not displayed for Translation Projects and QA Projects.

**Exclusion Settings** – during the extraction process, SDL MultiTerm Extract ignores any terms that already exist in the termbase and/or file that is specified under **Exclusion Settings**.

**Learning Settings** – during the extraction process, SDL MultiTerm Extract records any terms that you manually reject in a dedicated database. SDL MultiTerm Extract learns to automatically ignore these unwanted terms as extraction continues.

To specify the **Exclusion Settings**:

- Click **Exclude termbase**. The **Choose Exclude Termbase** dialog box is displayed.
From the **Choose Exclude Termbase** dialog box, select the termbase whose terms you want to exclude from the term extraction process.

- To select a local termbase, select the termbase and click **OK**. The **Choose Index to Exclude** dialog box is displayed.

- To select a remote termbase, click **Settings** to open the **Termbase Connection** dialog box. For more information about termbase connection settings, go to step 6 on page 3-9. After returning to the **Choose Exclude Termbase** dialog box, select the termbase and click **OK**. The **Choose Index to Exclude** dialog box is displayed.

**NOTE**

If you selected a project termbase already in step 5, this is automatically displayed under **Exclusion Settings**.

- Complete the **Choose Index to Exclude** dialog box.

- Select the source index to which SDL MultiTerm Extract should refer when checking for terms to exclude from the term extraction process.

- Click **OK** to return to the **Excluded terms** page.
On the **Excluded terms** page of the **New Project Wizard**, click **Exclude file** to specify lists of terms you do not want to include in the extraction process. The **Select** dialog box is displayed.

![Select dialog box](image)

- Browse to the exclude list and choose the list.

**NOTE**

SDL MultiTerm Extract comes with basic vocabulary lists for certain European languages. If there is a basic vocabulary list available for your source language, this is automatically displayed.

- Click **Select** to return to the **Excluded terms** page.

To specify the **Learning Settings**:

- On the **Excluded terms** page of the **New Project Wizard**, select the **Record discarded terms** check box to choose a learning database.

- Click **Browse** to select the database or to create a new one and the **Choose Learning Termbase** is displayed.

![Choose Learning Termbase](image)
If you want to select a remote termbase, click **Settings** to open the **Termbase Connection** dialog box. For more information about termbase connection settings, go to step 6 on page 3-9.

To create a new termbase, click **New termbase** and the **Create Learning Termbase** dialog box is displayed. Enter a **Termbase name** and select the termbase languages. Click **Create** and **OK** to return to the **Excluded terms** page.

Click **Next**. Depending on the type of project you are creating, one of three pages is displayed. Use these pages to customize the SDL MultiTerm Extract settings.

- For **Monolingual Term Extraction Projects** and **Bilingual Term Extraction Projects**, go to step 11 on page 3-16.
- For **Translation Projects** and **QA Projects**, go to step 12 on page 3-17.
- For **Dictionary Compilation Projects**, go to step 13 on page 3-19.
For **Monolingual Term Extraction Projects** and **Bilingual Term Extraction Projects**, the **Term extraction settings** page is displayed. Use this page to customize the SDL MultiTerm Extract term extraction settings.

- In the **Minimum term length** box, specify the minimum number of words required to form a term candidate. The default setting is 1.
- In the **Maximum term length** box, specify the maximum number of words that a term candidate may have. The default setting is 10.
- Select **Maximum number of extracted terms** to set the maximum number of term candidates that SDL MultiTerm Extract will extract. The default setting is 100.
- Set the **Silence/noise ratio**. The higher the silence ratio, the fewer term candidates are extracted. However, these term candidates are generally of a higher quality. The higher the noise ratio, the more term candidates are extracted. However, it is likely that these will include a number of lower quality candidates.
- Click **Stopword Lists** to set the language-specific list(s) containing articles, pronouns and other items that are to be excluded from the extraction process.
For **Translation Projects** and **QA Projects**, the **Term Recognition Options** page is displayed. The options available are different depending on if a **Translation Project** or **QA Project** was selected. Use these pages to customize the SDL MultiTerm Extract term recognition settings.

The following page is the **Term Recognition Options** for a **Translation Project**.

- Select **Only show terms in project termbase that are found in documents** to only include terms from the termbase that are displayed in the project file (**Translation Projects** only).
- Select **Show all terms in project termbase** to include all terms in the project termbase (**Translation Projects** only).
The following page displays the **Term Recognition Options** for a **QA Project**.

**NOTE**

The options available on this page are relevant to both **Translation Projects** and **QA Projects**.

- Use the **Minimum Match Value %** box to specify a value below which terms will not be included in the project.

- Use the **Search-Depth (10-999)** box to specify the number of terms to search at a time. When scanning source segments for terminology, SDL MultiTerm Extract evaluates not only single words but also multiple word units, trying to find similar terms in the current termbase. The search depth value defines the number of fuzzy matches SDL MultiTerm Extract should return. Increasing the **Search-Depth** value slows down terminology recognition. You can set it to any value between 10 and 999; however, high values are only necessary if you are working with large terminology databases. The default **Search-Depth** setting is 20 (matches) and in most cases this should produce satisfactory results.
For **Dictionary Compilation Projects**, the **Dictionary compilation settings** page is displayed. Use this page to customize the SDL MultiTerm Extract dictionary compilation settings.

- **Algorithm**: 
  - Minimum word frequency: 3
  - Maximum word frequency: 20
  - Maximum number of translations: 5
  - Minimum translation frequency: 3
  - Maximum translation frequency: 20

- **Stopword Lists**: to select the language-specific list(s) containing articles, pronouns and other items that are to be excluded from the extraction process.

- In the **Minimum word frequency** box, specify the minimum number of times a term and its translation must be shown in a text before being added to the term candidate list. The default setting is 3.

- Select the **Maximum word frequency** check box to specify the maximum number of times a term and its translation can be shown in a text before being excluded from the term candidate list. The default setting is 20.

- Select the **Maximum number of translations** check box to specify the maximum number of possible translations that SDL MultiTerm Extract will extract for a given term. This option is active by default and the default setting is 5.

- In the **Minimum translation frequency** box, specify the minimum number of times a term must be translated by its target term in order for the target term to be added to the term candidate list. The default setting is 3.

- Select the **Maximum translation frequency** check box to specify the maximum number of times a term may be translated by its target term before the target term is excluded from the term candidate list. The default setting is 20.

- Click **Stopword Lists** to select the language-specific list(s) containing articles, pronouns and other items that are to be excluded from the extraction process.
14 Click **Next**. For **Bilingual Term Extraction Projects** and **Translation Projects**, the **Translation settings** page is displayed. For all other projects, go to step 15 on page 3-21.

Use this page to customize the SDL MultiTerm Extract translation settings.

Select the **Search for new translations** check box to enable SDL MultiTerm Extract to extract translations and to specify the **Translation Options**.

- Select the **Maximum number of translations** check box to set the maximum number of possible translations that SDL MultiTerm Extract will extract for a given term. By default, this check box is selected and the default setting is 5.

- In the **Minimum translation frequency** box, specify the minimum number of times a term must be translated by a certain target term for the target term to be added to the term candidate list. The default setting is 3.

- Click **Stopword Lists** to set the language-specific list(s) containing articles, pronouns and other items that are to be excluded from the extraction process.
15 Click **Next**. The **Completing the New Project Wizard** page is displayed. Use this page to review the settings you specified for the created project.

![Completing the New Project Wizard](image)

**New Project Wizard**

**Completing the New Project Wizard**

You have successfully completed all the steps in the New Project Wizard.

Settings Summary:

- **Type**: Bilingual Term Extraction Project
- **Languages**: German(de) - English(en_GB)

Click **Finish** to create the project.

16 Click **Finish**. An SDL MultiTerm Extract **Confirm** dialog box is displayed.

![Confirm](image)

- Click **Yes** to start processing the project.
- Click **No** to return to the SDL MultiTerm Extract interface.

**SAVING A PROJECT**

Once a project is created, we recommend saving changes to your SDL MultiTerm Extract project regularly. There are three methods to save changes to your project:

- Select **File > Save Project** or **File > Save Project As** from the menu bar.
- Use the [Ctrl]+[S] keyboard combination.
- Click **Save Project** on the SDL MultiTerm Extract file toolbar.
IMPORTING TERMS INTO TRANSLATION PROJECTS

There may be times when you would like to add external terms to your SDL MultiTerm Extract Translation Project. An efficient way to add the terms is from a text file as described below.

To import terms from a text file:

1. Select Import from the File menu. The Import dialog box is displayed.

2. Specify the name and location of the file from which you want to import terms in the Import file box, or click Browse to locate the file.

3. Specify the type of encoding in the file from the Encoding drop-down list. The encoding available depends on the source language selected for the project. For example, you may have a choice of ANSI, Unicode or UTF8 text.

4. Click OK to begin the import.

5. The Import dialog box displays a progress bar. When the import process is complete, click OK to return to SDL MultiTerm Extract.
EXTRACTING TERMS

After a project is created, the next step is to extract the terms from the files you added during the project creation. Those terms are then added to a candidate term list in the project, for review and validation. Terms can be automatically or manually extracted from your project files.

Automatically Extracting Terms

There are two ways of starting the automatic term extraction process:

- SDL MultiTerm Extract prompts you to start the process after you add files to a project.
- Select Project > Process from the menu bar. You can also press [F7].

NOTE

The Process option is enabled after you create a new, or open an existing, SDL MultiTerm Extract project. The option is unavailable if no SDL MultiTerm Extract project is open or if no files have been added to the project.

The Term Extraction dialog box is displayed. A progress bar indicates the completion of the extraction process. The displayed actions performed, for example, Translating terms and Gathering QA statistics vary depending on the chosen project type.

Once the progress bar has reached 100%, click OK to view the extracted terms in SDL MultiTerm Extract.
Manually Extracting Terms

SDL MultiTerm Extract allows you to add more terms to the term candidate list by manually extracting terms from a document:

1. Open your SDL MultiTerm Extract project and select the source and target languages.
2. Select Project > Settings from the menu bar to open the Project Settings dialog box.
3. On the Add/Remove Files tab, click Browse to add the files from which to extract terms.
4. Click OK to close the dialog box.
5. Select View > Text Windows from the menu bar and select the document to display. The document opens in a text window, as shown below.

6. Newly extracted terms and translations are formatted by default in red text. Existing terms and translations are formatted by default in green text. SDL MultiTerm Extract assigns a line number to each sentence in the file. To add terms manually, select the text to add and do one of the following:
   - Select the text in the text window and click Add term or translation on the toolbar.
   - Press the keyboard shortcut [Ctrl]+[Insert].

Add term pairs by first selecting the source term. Next, right-click and choose Add as New Term from the shortcut menu. Alternatively, click Add term or translation on the toolbar. Select the corresponding translation, right-click and choose Add as Translation from the shortcut menu. Alternatively, click Add term or translation again on the toolbar.
Text Window

There are a number of options available in the text window toolbar:

- **Navigation**: Use Next page, Previous page, First page, Last page to scroll through the pages of the document in the text window.

- **Concordance**: Click Show concordance to display further concordance matches for the term candidate in the Concordance window.

- **Term properties**: Click Show term properties to display the term candidate in the Term properties window.

- **Go to**: Click Go to sentence to locate a specific sentence number by entering a number into the Go to sentence dialog box.

- Add terms by selecting the term, right-clicking and choosing Add as New Term from the shortcut menu or clicking Add term or translation.

- Remove terms by selecting the term, right-clicking and choosing Delete Term from the shortcut menu, clicking Delete term or translation or by pressing [Delete].

- The Show/Hide flag displays or hides the corresponding source and/or target language flags.

- Use Undo and Redo to undo or redo your most recent action.

- Click Help to open the online Help for SDL MultiTerm Extract.
This section describes how SDL MultiTerm Extract terms are displayed for viewing and validation.

The SDL MultiTerm Extract interface is divided into four parts:

- **Menu and toolbars** – Located at the top of the interface.
- **Term** window – Displays the extracted terms in spreadsheet format. If the project is a bilingual project the **Term** window also contains a list of translated terms.
- **Term properties** window – Displays additional information about a selected term and allows you to carry out global operations on the list of extracted terms.
- **Concordance** window – Displays the terms in context.

The next section explains how to use the **Term**, **Term properties** and **Concordance** windows to review and validate terms before exporting them.
Term Window

Extracted terms are displayed in the Term window.

The Term window contains a list of all the terms that SDL MultiTerm Extract has extracted in the current project. The spreadsheet presentation includes the following information about each term:

- **Filter**: Displays the active filter.
- **Find**: Search for candidate terms among the source terms in the Term window. This is not case-sensitive. After entering the text and viewing the results, you can return to the term list in the Term Window by deleting the string and pressing [Enter].
- **Score**: Displays the score that SDL MultiTerm Extract assigns to a term during the extraction process. The score is largely calculated using the frequency with which a term appears in the document. High scores indicate probable terms. This column is not available in the Translation Projects and Dictionary Compilation Projects.
- **Domain**: Shows the term category. You can change the domain in the Term window in the following ways:
  - For a selected term: right-click the selected term and choose a domain label from the Set Domain submenu on the shortcut menu. Alternatively, double-click the Domain column and choose a domain label from the drop-down list.
  - For several terms: select several terms at once by pressing [Ctrl]+[Shift] and clicking the relevant terms. Right-click any of the selected terms and choose a domain label from the Set Domain submenu on the shortcut menu.
NOTE

- Click **Settings** in the **Term properties** window to display the **Term Properties Window Options** dialog box. Use this dialog box to define the domain labels for your project. Once you have added a domain label there, it becomes automatically available from the **Domain** column in the **Term** window. See “Term Properties Window” on page 3-29.

- If the **Domain** column is hidden, right-click in the spreadsheet area or in the title bar of the **Term** window and select **Show Domain**. To hide it again, right-click in the spreadsheet area or in the spreadsheet header and select **Hide Domain**.

- Source Language Term: SDL MultiTerm Extract lists the extracted source language terms under the source language. If the source language term check box is selected, it indicates it is a validated term. When you first run the extraction process, none of the terms are validated.

- Translations: The translation(s) for the extracted term. SDL MultiTerm Extract lists translations under the target language. Translations also require validation. You can add a translation in the **Term** window by right-clicking and selecting **Add Translation** from the shortcut menu or by pressing [Ctrl]+[Insert], but this is only possible when no **Filter** is active.

NOTE

If an acronym is found, the text is placed in the corresponding box in the **Term properties** window. The definition or long form of the acronym is displayed in the window.

**Sorting Terms**

The default sort order for terms is by **Score** which lists terms starting with the most probable terms. You can also sort terms by domain or source language by clicking on the column label. A gray arrow indicates the current sort order.

**Deleting Terms**

To delete terms in the **Term** window, click to select the term to delete, right-click and choose **Delete** from the shortcut menu or press [Delete]. The term is removed from the term candidate list.
Term Properties Window

The Term properties window is divided into four areas:

- Global operation buttons
- Term and translation information
- Definition box and Note box
- Context information

Global Operation Buttons

- Click ◀ Go to first term, ▶ Go to last term, ◀ Go to previous term and ▶ Go to next term to scroll through the extracted terms and view their properties.
- Click   New term to add a new term to the extracted list.
- Click × Remove entry to remove the currently selected entry from the SDL MultiTerm Extract project.
- Click   Concordance to display concordance information about a term.
- Click   Settings to display the Term Properties Window Options dialog box. You can create new domains, specify the default domain and select which boxes are displayed in the Term properties window.
Term and Translation Information

This area displays general information about the selected term and its translations.

- **Source Language Term**: The source language term candidate as extracted by SDL MultiTerm Extract. The language is shown with a flag. Checked boxes indicate validated terms. If you have run the extraction process for the first time none of the terms will be validated.

- **Translation**: The target language is shown with a flag. You can validate a candidate translation by selecting the check box before the translation, or enter a new translation. To enter a new translation, select the **Click here to enter a new translation** checkbox and start typing. To delete a translation, select the term in the **Translation** box of the **Term properties** window and press [Delete]. Alternatively, select the term in the **Term** window and click **Delete**.

- **Domain**: This shows the term category. You can specify the domain for a selected term by choosing the domain from the drop-down list in the **Domain** box. Alternatively, you can click in the **Domain** box and enter a new domain label. The new label is added to the list of available domains.

- **Synonyms**: You can store synonyms of the term in this box. For example, you could enter lecture, discourse or oration.

- **Acronym**: You can store any acronym associated with the source term in this box.

- **Word forms**: Displays the word forms for the source term as found during term extraction. This box is read-only.

- **Filename**: Displays the name of the file(s) from which the source term was extracted. This box is read-only.

- **Created**: The date and time that the term was created. This information is displayed in the status bar.

- **Modified**: The date and time that the term was modified. This information is displayed in the status bar.

**NOTE**

When you perform an export, only the term and not the word forms are exported.

Definition Box

Add a definition of the term in the **Definition** box.
Note Box

Add any comments about a term or its translation(s) in the Note box.

Context Information

The Context sentences box stores sentence(s) showing the term in context.

- Click Add new sentence to enter a sentence that shows the term in context. Enter the sentence in the Context sentences box. The corresponding flags for each language are displayed before each sentence. Alternatively, you can right-click a sentence in the Concordance window and select Add As Context Sentence from the shortcut menu.

- Click Remove sentence to delete context sentences that are not suitable.

- Click Generate sentences to generate context sentences for a selected term or all terms. The default number of context sentences is five. Review and edit the displayed context sentences.

When you export terms, you can attach context sentences to each exported term as well as the name of the file from which the sentence was extracted.

NOTE

By default, terms and translations that were validated and exported from the project during previous edit sessions are formatted in green. You cannot edit these terms and translations.
Concordance Window

The Concordance window displays the extracted terms in context.

There are several ways of displaying a term in the Concordance window:

- Highlight the term in the Term window and press [Enter] to display the term in the Concordance window.
- Enter the term in the term box in the Concordance window and either press [Enter] or click Search. SDL MultiTerm Extract locates the term and displays it in the Concordance window.
- In the Term and text windows, select a term, right-click and choose Concordance from the shortcut menu.
- Click Concordance in the Term properties window.

NOTE
You can also select a term or place the cursor on it in the text window and press [F3] or click Show concordance.
If there is more than one concordance hit shown in the **Concordance** window, you can move from term to term using the following keyboard shortcuts:

- When you place the cursor on a term, the text changes to a blue color.
- Press `[Tab]` to go to the next occurrence of the term in the **Concordance** window.
- Press `[Shift]+[Tab]` to return to the previous occurrence of the term in the **Concordance** window.

Word forms and translations are displayed in context, that is, within a sentence. This is a sentence from the file from which they were extracted. The word form or translation is highlighted in the text using a different color, size or font, depending on the settings that you selected in the **Format** dialog box. You can access the **Format** dialog box by selecting **Tools > Format** from the menu bar. The selected term is displayed in blue. If you are working with a **Monolingual Term Extraction Project**, only the source language box is displayed.

### Adding or Deleting Terms in the Concordance Window

- To add a new term or translation from the **Concordance** window, select a term, right-click and choose **Add as New Term** or **Add as Translation** from the shortcut menu or press `[Ctrl]+[Insert]`. The term is displayed on the term candidate list immediately and is assigned a term score of 100.
- To delete a term or translation from the **Concordance** window, right-click and choose **Delete Term** if it is a source language term or **Delete Translation** if it is a target language translation from the shortcut menu or press `[Delete]`. The term is removed from the term candidate list.

### Formatting Terms in the Concordance Window

In the **Concordance** window, the default formatting for the text is Arial 12 point and Arial 12 point red for the highlighted word form or translation.
To modify the display of text and terms in the Concordance window:

1. Select Tools > Format from the menu bar. The Format dialog box is displayed.

   When the Text category is selected, the Preview pane shows how text will be displayed in the text window that is used for manual term extraction and in the Term, Term properties and Concordance windows.

   When the Concordance highlighting category is selected, the Preview pane shows how matching terms will be displayed in the text window that is used for manual term extraction and in the Concordance window.

   When the Existing Terms category is selected, the Preview pane shows how existing terms will be displayed in the Term, Term properties and Concordance windows.

2. Click Change to open the Choose Font dialog box.

   Select the Font, Font Colour, Size and Background Colour. Click OK to preview the font in the Format dialog box.

3. Click OK to apply the changes and to return to SDL MultiTerm Extract.

Repeat this process for each language in your SDL MultiTerm Extract Project.
VALIDATING EXTRACTED TERMS

To validate terms:

1. Review the term candidates in the Term window. Validate terms by selecting a term and pressing the space bar or selecting the check box. You can also right-click and select Validate from the shortcut menu. Use the down arrow to scroll through the term candidate list. You can delete terms that are not relevant.

2. After you review all term candidates in the Term window, use the Term properties window to examine terms in more detail. You can add new terms and remove term candidates that are not useful. In the Term window, select a term to display it in the Term properties window.

3. Review the term and its translations, using the Generate sentences button to refer to concordance sentences where necessary. Make any modifications required to the term properties.

4. Click Next term to display the next term in the list. Repeat steps 2 and 3.
5 Continue until you have reviewed and validated all the terms in the SDL MultiTerm Extract project.

NOTE
You can validate all term candidates by right-clicking in the Term window and selecting Validate All from the shortcut menu. Alternatively, to start the term review again, select Invalidate All to invalidate all term candidates.
This chapter provides information about how to export terms into other formats using Statistical Extraction in SDL MultiTerm Extract. Sections include:

- About Exporting Extracted Terms
- Exporting Using Export Definitions
- Exporting Suspect Sentences in a QA Project
- Using Filters when Exporting
ABOUT EXPORTING EXTRACTED TERMS

After your extracted terms have been validated in SDL MultiTerm Extract, they can be exported using the Export Wizard. This wizard takes you step-by-step through the process of exporting extracted terms from your SDL MultiTerm Extract project. You can export to a:

- Project Termbase
- MultiTerm Termbase
- MultiTerm XML Format
- Tab-delimited Text Format.

NOTE

If you are exporting to MultiTerm export format, SDL MultiTerm Extract allows you to add extracted terms to your existing termbase. You can also exclude terms from being added to your existing termbase by the term extraction process. You should close your termbase before starting the export.

Export Definitions

During the export process an export definition file is used containing information such as the export type and file format. When exporting you can use an existing export definition file or create a new one. MultiTerm is the default export type when creating a new export definition file.

Before You Begin

If you intend to export data directly to a remote MultiTerm termbase, the following information is required to define termbase connection settings:

- MultiTerm Server computer details – determine the name of the MultiTerm Server computer that provides access to the termbases with which you want to work.

- MultiTerm Server connection type – determine if your computer and the MultiTerm Server computer are linked by a standard LAN or a MultiTerm Anywhere connection. A standard LAN connection applies when both computers belong to the same physical domain. A MultiTerm Anywhere connection applies when the two computers belong to separate physical domains.
MultiTerm Server logon details – determine the user name and password of your MultiTerm Server login.

Contact your MultiTerm administrator for these details and for information about your termbase user rights. You must have write permission in the appropriate areas of the specified remote termbase to export data directly from SDL MultiTerm Extract. If you intend to export directly into a local termbase, determine the name and location before you start the Export Wizard. You have exclusive access and unlimited user rights in all local termbases, and no login details are required.

EXPORTING TERMS

To export extracted terms from SDL MultiTerm Extract:

1. Select File > Export from the menu bar. This Export Wizard is displayed.

2. Click Next. The Export Definition page is displayed. Use this page to select the export type.
To export to a project termbase, select Export to the project termbase.

To export to a MultiTerm termbase, select Create a new export definition and select MultiTerm 7 from the drop-down list.

To export to MultiTerm XML format, select Create a new export definition and select MultiTerm XML Format.

To export to tab-delimited text format, select Create a new export definition and select Tab delimited from the drop-down list.

To export using an existing export definition, select Export using an existing export definition. See “Exporting Using Export Definitions” on page 4-11.

3 Click Next. The Export Settings page is displayed. The options on this page are different depending on which type of export you selected. Use this page to define the export settings.

To export to a project termbase, this page is not displayed, go to step 7.

To export to a MultiTerm termbase, go to step 4.

To export to MultiTerm XML format, go to step 5.

To export to tab-delimited text format, go to step 6.

4 If you selected to export to a MultiTerm termbase, the following page is displayed.

- Select Export to termbase. Click Browse to locate the termbase.

NOTE
You may be required to reconnect to the termbase. If this happens, you must log on again.
5 If you selected to export to MultiTerm XML format, the following page is displayed.

- Select the Export File Location. By default, SDL MultiTerm Extract selects the current projects folder and creates an XML file (*.xml) with the same name as the project name.

- Specify a Termbase or termbase definition File (*.xdt). The selected item provides SDL MultiTerm Extract with the termbase definition reference that is required to create the export file.

6 If you selected to export to tab-delimited text format, the following page is displayed.

- Select the Export File Location. By default, SDL MultiTerm Extract selects the current projects folder and creates a text file (*.txt) with the same name as the project name.
Click **Next**. The **Filters** page is displayed. Use this page to select or create a new filter.

You can select a filter from the drop-down list. SDL MultiTerm Extract ships with two predefined filter definitions: **Only Validated** and **Only Non Validated** terms. The default setting is **No Filter**.

**NOTE**

Click **Create New Filter** to create a new filter or click **Edit Filter** to edit an existing filter. See “Using Filters when Exporting” on page 4-16.

Click **Next**. The **Export Languages** page is displayed. Use this page to select the languages you want to export.

This page is not displayed if you selected to export to tab-delimited text format, go to step 13 on page 4-9.

On the **Export Languages** page, map the languages in your SDL MultiTerm Extract project to the corresponding index in the MultiTerm termbase, project termbase or MultiTerm XML format. Select a MultiTerm index from the drop-down list for each language.
9 Click **Advanced**. The **Export Schema** dialog box is displayed. Use this page to create an export schema.

SDL MultiTerm Extract automatically exports source and target terms to the termbase. If you create an export schema, you can also export the following information to the selected termbase: **Domain**, **Definition** and **Note** boxes. For source and/or target languages you can export Validation status information (**Validated**), **Score**, **Context** (and if required, context sentences and context sentence file name), **File name**, **Acronym**, **Modified** and **Created** boxes.

10 In the **Exported Field** box, select a checkbox to export information for that item. For example, select **Context** to export **Context** information. The boxes pertaining to that selected checkbox are displayed on the right side of the dialog box.
In the Level box, select a level from the drop-down list to indicate which level the term is to be exported to the termbase. Each box should be associated with a level in MultiTerm, for example, Entry level fields, Index level fields and Term level fields.

Select the Field in the MultiTerm termbase to which to export Context from the drop-down list. SDL MultiTerm Extract displays the MultiTerm termbase fields in a drop-down list.

Select Automatically generate context sentences to add context information to the termbase. The default setting for this option is 10 sentences. If you want to change the number of sentences, enter a different number.

Select Export context sentence filename to export this information.

11 Click OK to return to the Export Wizard.

12 Click Next. The Termbase Export Options page is displayed. Use this page to select different export options.

**NOTE**

This page is only displayed for exporting to a MultiTerm termbase or project termbase. For all other projects, go to step 13 on page 4-9.

Select Export all terms as new entries if each term is to create a new entry. You can choose between Export all translations to one new entry, if all translations are to be exported under the same entry or Export each translation to a separate new entry, if each translation of a given source term is to be exported under a different entry.

Select Update existing concepts and merge on source index term if all termbase entries which have the same index term are to be merged and updated with newly exported information.

Select Prompt for each term to determine how each term is exported.
13 Click **Next**. The **Completing the Export Wizard** page is displayed. Use this page to review a summary of the options you selected for this export.

![Completing the Export Wizard](image)

Click **Finish** to start the export process.

14 SDL MultiTerm Extract prompts you to save the export definition for future use. Click **Yes** if you want to save the export definition for re-use. For more information, See “Exporting Using Export Definitions” on page 4-11.

![Confirm](image)

15 The **Choose Export Definition** dialog box opens. Enter the file name for the export definition and click **Set**.

![Choose Export Definition](image)
16 If you are exporting to a remote termbase, a message is displayed asking you to specify the termbase entry class to which data will be exported. Select the entry class from the drop-down list.

![Select Entry Class dialog box]

NOTE
You must have write permission for the specified termbase entry class to export data directly from SDL MultiTerm Extract. If necessary, contact your MultiTerm administrator for precise details of your termbase user rights.

17 The Export dialog box is displayed. A progress bar indicates the export status. When the progress bar reaches 100%, click OK to exit the Export Wizard and return to SDL MultiTerm Extract.
EXPORTING USING EXPORT DEFINITIONS

An export definition file defines the export type and file format for export. The options displayed during the Export Wizard vary depending on which export definition file is selected. You can use an existing export definition file or create a new export definition file when exporting.

Exporting Using an Existing Export Definition

To export extracted terms from SDL MultiTerm Extract using an existing export definition:

1. Select File > Export from the menu bar. This the Export Wizard is displayed.

2. Click Next. The Export Definition page is displayed. Use this page to select the export type.

3. Select Export using an existing export definition and click Browse to select SDL MultiTerm Extract export definition file (*.exd). Click Set.
4 The **Export Wizard** guides you through the steps required for the chosen export definition.

5 When you reach the final page in the wizard, **Completing the Export Wizard**, click **Finish** to start the export process and to close the **Export Wizard**.

---

**Exporting Using a New Export Definition**

To export extracted terms from SDL MultiTerm Extract using a new export definition file:

1 Select **File > Export** from the menu bar. This **Export Wizard** is displayed.
2. Click **Next**. The **Export Definition** page is displayed. Use this page to select the export type.

![Export Definition Page]

3. Select **Create a new export definition** and select the export type from the drop-down list.

4. Click **Next**. The **Export Wizard** guides you through the steps required for the new export definition. Options vary depending on which export format is selected. See “Exporting Terms” on page 4-3.

5. When you reach the final page in the wizard, **Completing the Export Wizard**, click **Finish** to start the export process and to close the **Export Wizard**.

![Completing the Export Wizard]

6. SDL MultiTerm Extract prompts you to save the export definition for future use. Click **Yes** to save the export definition for re-use.

![Save Export Definition]
7 The **Choose Export Definition** dialog box opens. Enter the file name for the export definition and click **Set**.

8 If you are exporting to a remote termbase, a message is displayed asking you to specify the termbase entry class to which data will be exported. Select the entry class from the drop-down list.

**NOTE**

You must have write permission for the specified termbase entry class to export data directly from SDL MultiTerm Extract. If necessary, contact your MultiTerm administrator for precise details of your termbase user rights.

9 The **Export** dialog box is displayed. A progress bar indicates the export status. When the progress bar reaches 100%, click **OK** to exit the **Export Wizard** and return to SDL MultiTerm Extract.
EXPORTING SUSPECT SENTENCES IN A QA PROJECT

You can export suspect sentences from a file. This procedure can only be performed in a QA Project. Suspect sentences are sentences that contain terminology that does not match a translation in the termbase. Exporting suspect sentences allows you to edit them as a group.

To export suspect sentences:

1. Select File > Export suspect sentences from the menu bar or press [Ctrl]+[E]. The Export Suspect Sentences dialog box is displayed.

2. Select the file from which you want to export suspect sentences.

3. Select sentences to export that contain the following:
   - Terms that don’t have a translation in the termbase
   - Terms that have a translation that is not in the termbase
   - All categories (all suspect sentences).

4. Use Browse to select an output file.

5. Click Export. The sentences are exported to the output file.

6. Open the output file and edit the sentences as required.

**NOTE**

In the text window in the QA Project, you can select from a drop-down list which category of suspect sentences you want to highlight. You can navigate through the terms by pressing [Tab] to go to the next term and [Shift]+[Tab] to go to the previous term. To move from one suspect sentence to the next press [Ctrl]+[Tab].
USING FILTERS WHEN EXPORTING

SDL MultiTerm Extract uses filters in the export process when determining which terms to export. This section explains the following tasks:

- Creating a new filter
- Editing a filter definition
- Removing a filter.

Creating a Filter

To create a filter:

1. Select Tools > Filters from the menu bar. The Filters dialog box opens. Click New to open the New Filter dialog box.

   ![New Filter dialog box](image)

   **TIP**
   
   You can also access the New Filter dialog box from the Export Wizard. Click Create New Filter or Edit Filter on the Filters page. For more information, go to step 7 on page 4-6.

2. In the Name box, enter a name for the filter to create. The default name for the filter is MyFilter.
To create a filter definition:

- Select a box from the Field List and click Add to add it to the Constraints List.

- Specify a criterion to define how SDL MultiTerm Extract should filter on the selected box. If the selected box is associated with an attribute picklist, select an item from the Attribute Picklist. If the selected box is not associated with an attribute picklist, use the Criterion box to enter a suitable criterion. The criterion you specify is automatically added to the Constraints List.

In the following screen shot, the filter definition consists of one constraint. The box Validated is added to the Constraints List and Yes is specified as the criterion. When this filter is active, SDL MultiTerm Extract will export validated terms only.

- Repeat this process for each constraint to add to the filter definition. A filter definition may consist of one or several constraints.

**NOTE**

The tree structure in the Field List indicates whether a box applies to entry level, source index or target index content. The level to which a box belongs determines whether the filter constraint is valid for entry level, source index or target index content.
The criteria that you can apply to the various boxes in the Field List are outlined below.

**Entry level boxes:**

- **Note**: Enter a text string to filter on in the Note box. If you select <Empty>, and choose NOT as the operator, SDL MultiTerm Extract filters for all terms that contain a note.

- **Definition**: Enter a text string to filter on in the Definition box. If you select <Empty>, and choose NOT as the operator, SDL MultiTerm Extract filters for all terms that contain a definition.

- **Domain**: Enter a domain label to filter on.

- **Number of translations**: Enter the greater than [>] or less than [<] or equal [=] sign and a value to indicate the number of translations. For example, if you specify < 10, SDL MultiTerm Extract will filter for all terms with fewer than 10 translations.

- **Created**: Enter the date on which the entry was created. Use the greater than [>] or less than [<] or equal [=] signs to qualify the criterion.

- **Modified**: Enter the date on which the entry was modified. Use the greater than [>] or less than [<] or equal [=] signs to qualify the criterion.

**Source index and target index boxes:**

- **Score**: Enter the greater than [>] or less than [<] or equal [=] sign and a value to indicate the score. For example, if you specify a score of > 70, SDL MultiTerm Extract will filter for all terms with a score that is greater than 70. The score is largely calculated using the frequency that a term appears in the document. High scores indicate probable terms.

- **Validated**: From the Attribute Picklist, select Yes to include and No to exclude validated terms.

- **File name**: Enter the name of the file on which you want to filter.

- **Acronym**: Enter the acronyms on which you want to filter; click Or to separate them with a logical 'or'. You can enter the full form of the acronym if the acronym is defined in the Acronym box in the Term properties window. If you select <Empty>, and choose NOT as the operator, SDL MultiTerm Extract filters for all terms that contain an acronym.

- **Existing**: Select Yes to include terms that already exist in the termbase and No to exclude existing terms.
You can further qualify your filter constraints as follows:

- Select **Case sensitive** for the text strings you enter as filter criteria to be case-sensitive. For example, if you enter the text string *ansi* as a filter criteria and select the **Case sensitive** option, SDL MultiTerm Extract will filter on *ansi* but not on *Ansi* or *ANSI*.

- To separate items on the **Constraints List** with a logical ‘or’, click **Or**.

- To exclude specific content from the export, click **Not** to qualify the relevant filter constraint.

4. When you have finished defining your filter, check that the information displayed in the **Constraints List** is correct.

5. Click **OK** to confirm and to return to the **Filters** dialog box. The new filter is displayed on the list and will be available for selection during the export process.

6. To create another filter, click **New**. To return to your SDL MultiTerm Extract project, click **OK**.

**Editing a Filter**

To edit a filter definition:

1. Select **Tools > Filters** from the menu bar. The **Filters** dialog box opens. Select the filter definition to modify and click **Edit** to open the **Edit Filter** dialog box.

   **TIP**

   You can also access the **New Filter** dialog box from the **Export Wizard**. To do this, click **Create New Filter** or **Edit Filter** on the **Filters** page. For more information on filters, go to step 7 on page 4-6.

2. Change the filter definition as required. Click **OK** to return to the **Filters** dialog box. Click **OK** again to return to your SDL MultiTerm Extract project.

**Removing a Filter**

To remove a filter definition:

1. Select **Tools > Filters** from the menu bar. The **Filters** dialog box opens. Select the filter definition to remove and click **Remove**.

2. The filter is removed from the list. Click **OK** to return to your SDL MultiTerm Extract project.
This chapter gives an overview and description of the components included in SDL MultiTerm Extract. Sections include:

- Settings
- QA Statistics
- Keyboard Shortcuts
SETTINGS

This section includes information on defining settings and preferences for SDL Trados projects. The following areas are presented:

- Preferences
- Project settings
- File Format Settings

Preferences

Select File > Preferences from the menu bar to define your SDL MultiTerm Extract project preferences. The Preferences dialog box opens.

- Select the Remove non-validated terms before extracting check box to be asked if SDL MultiTerm Extract should remove any non-validated terms from the term candidate list. This option applies when you process a project for a second or subsequent time within the same session. Non-validated terms are added to the learning termbase for your project.
- Select the Open most recent project at startup check box to launch the last project you worked on when you opened SDL MultiTerm Extract.
- Select the Ask for confirmation when exiting MultiTerm Extract check box to be prompted to confirm that you want to exit the application.
- Number of concordance sentences (on each page) determines how many concordance examples are displayed in one Concordance window. The default is set to 10.
- Click Reset to defaults to return to the SDL MultiTerm Extract default preference settings.
- Click OK when you are finished.
Project Settings

Select Project > Settings from the menu bar to open the Project Settings dialog box. SDL Trados collects all the information about the term extraction process. This section explains the tabs within the Project Settings dialog box. The tabs available depend on the project type and include the following:

- Add/Remove Files
- Exclude
- Term Extraction
- Dictionary Compilation
- Translation

Add/Remove Files Tab

The Project Files list displays the files in the current SDL Trados project.

To add more files to the project, click Browse to select the files to add.

To remove files from the project, select the files to remove and click Remove. A confirmation message is displayed. You can also remove the terms belonging to the files from the project. Click Yes and the selected files are removed from the Project Files list. If you choose to remove
the terms as well, they are removed from the Term window. Click No to return to the Project Settings dialog box.

Exclude Tab

On the Exclude tab, select termbases and exclude lists for your project. Terms in the selected termbases and exclude lists are not extracted from the files specified in your project settings. If you want to add extracted terms to an existing MultiTerm termbase, you exclude terms from that termbase from the extraction process to avoid double entries and repeatedly validating the same terms. An exclude list is a text file containing specific terms to not include in the extraction process.

To exclude extraction of terms from a termbase in your project:

1. Click Exclude termbase. The Choose Exclude Termbase dialog box is displayed.

2. Select the termbase to exclude and click OK. The Choose Index to Exclude dialog box is displayed.

   NOTE

   If you select a remote termbase, click Settings to open the Termbase Connection dialog box where you can specify MultiTerm Server computer and login details.

3. Select the language of the terms to exclude. This will usually be the same language as the source language of your project.

4. Click OK to return to the Project Settings dialog box.
To set exclude lists:

1. Click **Exclude file** to specify lists of terms to not include in the extraction process. The **Select** dialog box is displayed.

2. Browse to the exclude list and click **Select** to return to the **Project Settings** dialog box.

![NOTE]
- View the excluded terms by clicking **View excluded terms**.
- Remove an exclude list by selecting a file and clicking **Remove**.

### Defining the Learning Settings

SDL Trados uses the learning termbase to collect all term candidates that you delete, and thus reject, as term candidates during the term extraction process. The learning termbase is updated each time you delete a term during term extraction; terms that exist in the learning termbase are not extracted again. In this way, SDL Trados avoids displaying unwanted terms in the **Term** window.

You can use the learning termbase in the current project and in subsequent projects as an exclude termbase. SDL Trados does not extract any terms that exist in the exclude termbase during term extraction.

To select the learning settings for your project:

1. Select the **Record discarded terms** check box to choose a database that will store any terms you reject.

2. Click **Browse** to select the database or to create a new one. The **Choose Learning Termbase** dialog box is displayed.

   - Select the termbase from the **Termbases** list and click **OK**.
   - Click **Settings** to specify the termbase location.

![NOTE]
- Specify the MultiTerm Server computer if you select a remote termbase.
Click **New termbase** to create a new learning termbase. The **Create Learning Termbase** dialog box is displayed.

- Enter the name of the termbase in the **Termbase name** box.
- Select the languages you require from the **Select languages** list box.
- Click **Create** to set up the new termbase.

3 Click **Options** to display the **Learning Termbase Options** dialog box.

- To update the learning termbase every time you delete a term candidate from a project, select the appropriate option.
- To receive a confirmation message if entries are being added to the learning termbase, choose the appropriate option.

4 Click **OK** to return to **Project Settings**.

**Term Extraction Tab**

You can customize the following term extraction algorithm settings to fine tune the extraction process.
- **Minimum term length** indicates the minimum number of words required to form a term. The default setting is 1.

- **Maximum term length** indicates the maximum number of words that a term may have. The default setting is 10.

- **Maximum number of extracted terms** indicates the maximum number of terms that SDL Trados will extract. The default setting is 100.

- **Silence/noise ratio:** The higher the silence ratio, the fewer term candidates are extracted. However, these term candidates are generally of higher quality. The higher the noise ratio, the more term candidates are extracted. However, it is likely that these will include a number of lower quality candidates.

**Dictionary Compilation Tab**

- In the **Minimum word frequency** box, specify the minimum number of times a term and its translation must be shown in a text before being added to the term candidate list. The default setting is 3.

- Select **Maximum word frequency** to specify the maximum number of times a term and its translation can appear in a text before being excluded from the term candidate list. The default setting is 20.

- Select **Maximum number of translations** to specify the maximum number of possible translations that SDL Trados will extract for a given term. This option is active by default and the default setting is 5.
In the **Minimum translation frequency** box, specify the minimum number of times a term must be translated by its target in order for the target term to be added to the term candidate list. The default setting is 3.

In the **Maximum translation frequency** box, specify the maximum number of times a term may be translated by its target term before the target term is excluded from the term candidate list. The default setting is 20.

Click **Stopword Lists** to set the language-specific list(s) containing articles, pronouns and other items that are to be excluded from the extraction process.

**Translation Tab**

You can customize the following translation algorithm settings to fine tune the translation process.

- Select the **Search for new translations** check box.
- Select **Maximum number of translations** to set the maximum number of possible translations that SDL Trados will extract for a given term. This option is active by default and the default setting is 5.
- In the **Minimum translation frequency** box, specify the minimum number of times a term must be translated by its target term in order for the target term to be added to the term candidate list. The default setting is 3.
- Click **Stopword Lists** to set the language-specific list(s) containing articles, pronouns and other items that are to be excluded from the extraction process.
File Format Settings

To extract terms from HTML, SGML, XML, ASP or JSP files, you must specify a settings file (*.ini) before starting the term extraction process. SDL MultiTerm Extract needs the settings file to interpret these files correctly and to distinguish between tags and text. If you do not have a settings file, you must create one using TagEditor or Translator’s Workbench.

FOR MORE INFORMATION

For more information about creating settings files, see the online Help for TagEditor or Translator’s Workbench.

Selecting a Settings File

To select a settings file:

1. Select Tools > File Format Settings from the menu bar. The File Format Settings dialog box is displayed.

   ![File Format Settings dialog box]

TIP

You can also access the File Format Settings dialog box from the New Project Wizard. On the Add files to this project page, click File Format Settings. The File Format Settings dialog box is displayed where you can select a specific settings file for the current project.

SDL Trados provides default settings files for several tagged formats, including *.html, *.htm, *.jsp, *.asp, *.aspx, *.sgml, *.sgm and *.xml files. The default settings files are listed in the Settings File column.
2. Under **File Type**, identify the file format for which you want to specify a settings file and then click in the corresponding cell in the **Settings File** column. The relevant cell in the **File Type** column is highlighted and a **Browse** button is displayed opposite in the **Settings File** column.

3. Click **Browse** to open the **Choose Settings File** dialog box where you can browse to select the settings file. Click **Set** to return to the **File Format Settings** dialog box. The path to the settings file is shown in the **Settings File** column.

4. Click **OK** to confirm and to return to the main program window.
QA STATISTICS

After you have processed a QA Project, a QA report provides statistical information about the terms in project files and existing terms in the MultiTerm termbase for the project. The statistics are related to the quality of the project files or translated texts and the quality of the MultiTerm termbase. In particular, the report contains a number of tables and pie charts that are explained in the following section.

Document Quality

The table refers to source sentences containing terms that also exist in the project termbase.

- **Number of sentences without terms**: Number of source sentences that do not contain any terms that are stored in the project termbase.

- **Number of correct sentences**: Number of correct source language sentences containing terms that also exist in the project termbase. A ‘correct’ sentence is one whose target language terms are identical to those in the termbase.

- **Number of incorrect sentences**: Number of incorrect source language sentences containing terms that also exist in the project termbase. An ‘incorrect’ sentence is one in which at least one of the target language terms is different to its equivalent in the termbase.

- **Total number of sentences**: The total number of source language sentences.

![Document Quality Table]

Termbase Quality

![Termbase Quality Table]
- **Occurrences of terms with a translation**: Number of source language terms for which the target language term is identical to the one in the termbase.

- **Occurrences of terms without a translation**: Number of source language terms for which the target language term is different to its equivalent in the termbase.

- **Occurrences of terms**: Number of source language terms that exist in both the text and the termbase.

**Termbase Quality**

The information in this table shows the extent to which terms in the translated text match the translations given in the project termbase.

- **Terms which are always translated correctly**: Number of source language terms whose target language terms are identical to those in the project termbase.

- **Terms which are never translated correctly**: Number of source language terms whose target language terms are never identical to those in the project termbase.

- **Terms which are sometimes translated correctly**: Number of source language terms whose target language terms are sometimes identical and sometimes different to those in the project termbase.

- **Terms**: The total number of source language terms.
# KEYBOARD SHORTCUTS

The following table shows the keyboard shortcuts for use with SDL Trados.

<table>
<thead>
<tr>
<th>Keyboard Shortcut</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Menus</strong></td>
<td></td>
</tr>
<tr>
<td>[Ctrl]+[N]</td>
<td>Creates a new project.</td>
</tr>
<tr>
<td>[Ctrl]+[O]</td>
<td>Opens an existing project.</td>
</tr>
<tr>
<td>[Ctrl]+[W]</td>
<td>Closes a project.</td>
</tr>
<tr>
<td>[Ctrl]+[S]</td>
<td>Saves a project.</td>
</tr>
<tr>
<td>[Ctrl]+[A]</td>
<td>Saves a project under a different name.</td>
</tr>
<tr>
<td>[Ctrl]+[E]</td>
<td>Exports terms. In a <strong>QA Project</strong> this exports suspect sentences.</td>
</tr>
<tr>
<td>[Ctrl]+[Z]</td>
<td>Undoes the last action.</td>
</tr>
<tr>
<td>[Ctrl]+[Y]</td>
<td>Redoes the last action.</td>
</tr>
<tr>
<td>F7</td>
<td>Processes files.</td>
</tr>
<tr>
<td><strong>Term Window</strong></td>
<td></td>
</tr>
<tr>
<td>[Spacebar]</td>
<td>Validates the current term candidate.</td>
</tr>
<tr>
<td>[Delete]</td>
<td>Deletes the selected term from the term candidates list.</td>
</tr>
<tr>
<td>[F2]</td>
<td>Opens a term in the source and/or target column for editing.</td>
</tr>
<tr>
<td>[Ctrl]+[Insert]</td>
<td>Inserts a new translation of a source term in the target column of the Term window.</td>
</tr>
<tr>
<td><strong>Text Window</strong></td>
<td></td>
</tr>
<tr>
<td>[Spacebar]</td>
<td>Selects the term candidate or translation at the cursor.</td>
</tr>
<tr>
<td>[Delete]</td>
<td>Deletes the current term if it is part of the term candidates list.</td>
</tr>
<tr>
<td>[Ctrl]+[Insert]</td>
<td>If source sentence text is selected, this command adds selected text in a source sentence as a new term. If target sentence text is selected, this command adds selected text in a target sentence as a new translation.</td>
</tr>
<tr>
<td>[F3]</td>
<td>If some text is selected, SDL Trados carries out a concordance search using this text.</td>
</tr>
<tr>
<td>[Tab]</td>
<td>In the <strong>QA Project</strong>, goes to the next term.</td>
</tr>
<tr>
<td>Shortcut</td>
<td>Function</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>[Shift]+[Tab]</td>
<td>In the QA project, goes to the previous term.</td>
</tr>
<tr>
<td>[Ctrl]+[Tab]</td>
<td>In the QA project, moves from one suspect sentence to the next.</td>
</tr>
<tr>
<td>[Alt]+[Enter]</td>
<td>Displays the properties of the current term in the Term properties window.</td>
</tr>
</tbody>
</table>

**Concordance Window**

<table>
<thead>
<tr>
<th>Key</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Spacebar]</td>
<td>Selects the term candidate or translation at the cursor.</td>
</tr>
<tr>
<td>[Delete]</td>
<td>Deletes the current term if it is part of the term candidates list.</td>
</tr>
<tr>
<td>[F3]</td>
<td>If some text is selected SDL Trados carries out a concordance search using this text.</td>
</tr>
<tr>
<td>[Ctrl]+[Insert]</td>
<td>Adds selected text as new term.</td>
</tr>
<tr>
<td>[Shift]+[Tab]</td>
<td>Goes to the previous term in the Concordance window.</td>
</tr>
<tr>
<td>[Tab]</td>
<td>Goes to the next term in the Concordance window.</td>
</tr>
<tr>
<td>[Alt]+[Enter]</td>
<td>Displays the properties of the current term in the Term properties window.</td>
</tr>
</tbody>
</table>

**Other Shortcuts**

<table>
<thead>
<tr>
<th>Key</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Ctrl]+[Tab]</td>
<td>Switches between internal frames in the Term, Term properties and Concordance windows.</td>
</tr>
</tbody>
</table>
SDL is the leader in Global Information Management. Global Information Management enables companies to engage with their customers throughout the customer journey – from brand awareness, to sales and after-sales support – and across languages, cultures and channels.

SDL’s best-of-breed Web Content Management, eCommerce, Structured Content and Language Technologies, combined with its Language Services drive down the cost of content creation, management, translation and publishing. SDL solutions increase conversion ratios and customer satisfaction through targeted information across all customer touch points.

Global industry leaders who rely on SDL include ABN-Amro, Bosch, Canon, CNH, FICO, GlaxoSmithKline, Hewlett-Packard, KLM, Microsoft, NetApp, Philips, SAP and Sony. SDL has over 1,500 enterprise customers, has deployed over 170,000 software licenses and provides access to on-demand portals for 10 million customers per month. It has a global infrastructure of more than 60 offices in 35 countries.

For more information, visit www.sdl.com.