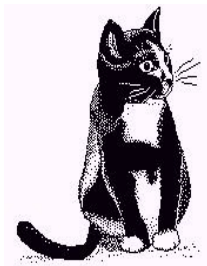


**OmegaT for**



**CAT**

**Beginners**

by Susan Welsh & Marc Prior

## Copyright

Copyright © 2014 Susan Welsh and Marc Prior

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License.”

Cover illustration is from [www.freeclipartnow.com](http://www.freeclipartnow.com), in the public domain.

Last updated: March 2014

Refers to OmegaT version: 3.0.8\_2

Screenshots from OmegaT versions: 1.6.0, 3.0.8\_2

Please note that owing to the pace at which OmegaT is being developed, the appearance of the screenshots and possibly some other information may have changed slightly.

## Table of Contents

Copyright.....	2
Introduction.....	5
Intended readership.....	5
What is a CAT tool and why are they useful?.....	5
1. Downloading OmegaT.....	5
2. Installing OmegaT.....	5
3. The OmegaT user interface.....	5
4. Installing a spelling dictionary.....	6
5. Creating a project.....	7
6. Translating.....	7
6.A. Project 1: A simple MS Word file.....	7
6.A.1. Handling tags.....	8
6.A.2. Checking your translation.....	8
6.A.3. Validating tags, creating target documents.....	8
6.A.4. Final proofreading, correction, and delivery.....	9
6.B. Project 2: An HTML document.....	9
6.B.1. Downloading the supplementary materials.....	9
6.B.2. Translation memory (fuzzy match function).....	18
6.B.3. The glossary function.....	19
6.B.4. The text search (find) function.....	19
7. The OmegaT user group.....	19
8. Appendix.....	19
GNU Free Documentation License.....	19
0. Preamble.....	19
1. APPLICABILITY AND DEFINITIONS.....	20
2. VERBATIM COPYING.....	20
3. COPYING IN QUANTITY.....	20
4. MODIFICATIONS.....	20
5. COMBINING DOCUMENTS.....	21
6. COLLECTIONS OF DOCUMENTS.....	22
7. AGGREGATION WITH INDEPENDENT WORKS.....	22
8. TRANSLATION.....	22
9. TERMINATION.....	22
10. FUTURE REVISIONS OF THIS LICENSE.....	22



# Introduction

## Intended readership

This document is intended for translators who are not experienced with Computer Assisted Translation (CAT) tools, and discusses only the basics. Another useful source of information is [www.omegat.org/en/documentation.php](http://www.omegat.org/en/documentation.php).

## What is a CAT tool and why are they useful?

A CAT tool, as the term is generally used by translators, refers to a software program that creates a “translation memory” (TM) of a translated document, which can then be used to facilitate the translation of future documents that have significant similarity to the first. This aspect makes it especially useful for repetitive material. Also:

- A glossary function allows the user to enter source words and their translations into a glossary, which can be accessed while translating.
- A search function scans previously translated material (either in a document in the present project, or in a translation memory from a previous project), so that the user can see how a word or phrase was previously translated.
- A segmentation function displays one “segment” of source text at a time (usually a sentence), and the translation is done with the target text on the screen directly below the source text. This is very useful for 1) making sure you don’t skip text; 2) checking your work; and 3) putting both source and target text in a font and size that are comfortable for you.

CAT tools allow for greater consistency of translation, including when more than one translator is working on a job. Finally, CAT tools reproduce the formatting of the source document. CAT tools range in price from free (OmegaT) to \$2,500 or more. OmegaT is an excellent place to start, as the price is right, it is relatively easy to learn, and it runs on MAC and under Linux, as well as Windows.

OmegaT supports the current Microsoft Office file formats (.docx, .xlsx, .pptx). It does not support the older Microsoft Office file formats (.doc, .xls, .ppt); these must first be converted, e.g., to the current Microsoft Office file formats.

See the following for more information on CAT tools, including on the many varieties available:

[http://en.wikipedia.org/wiki/Computer-assisted\\_translation](http://en.wikipedia.org/wiki/Computer-assisted_translation)

## 1. Downloading OmegaT

Download OmegaT (the OmegaT site tells you the various options; “beta” versions are stable, but may have a bug or two and do not have up-to-date documentation).

[www.omegat.org](http://www.omegat.org)

The on-screen download instructions should be self-explanatory. It is advisable to choose a version that includes Java (JRE) in your download. For the purposes of this tutorial, we assume that you will be using a Windows operating system. If you are using another system, and run into problems, the OmegaT users group at Yahoo can help you (see Step **8**).

## **2. Installing OmegaT**

When/after downloading the zip file, place it in a suitable folder, e.g., for Windows, C:\Program Files.

Unpack (extract, unzip) the zip archive.

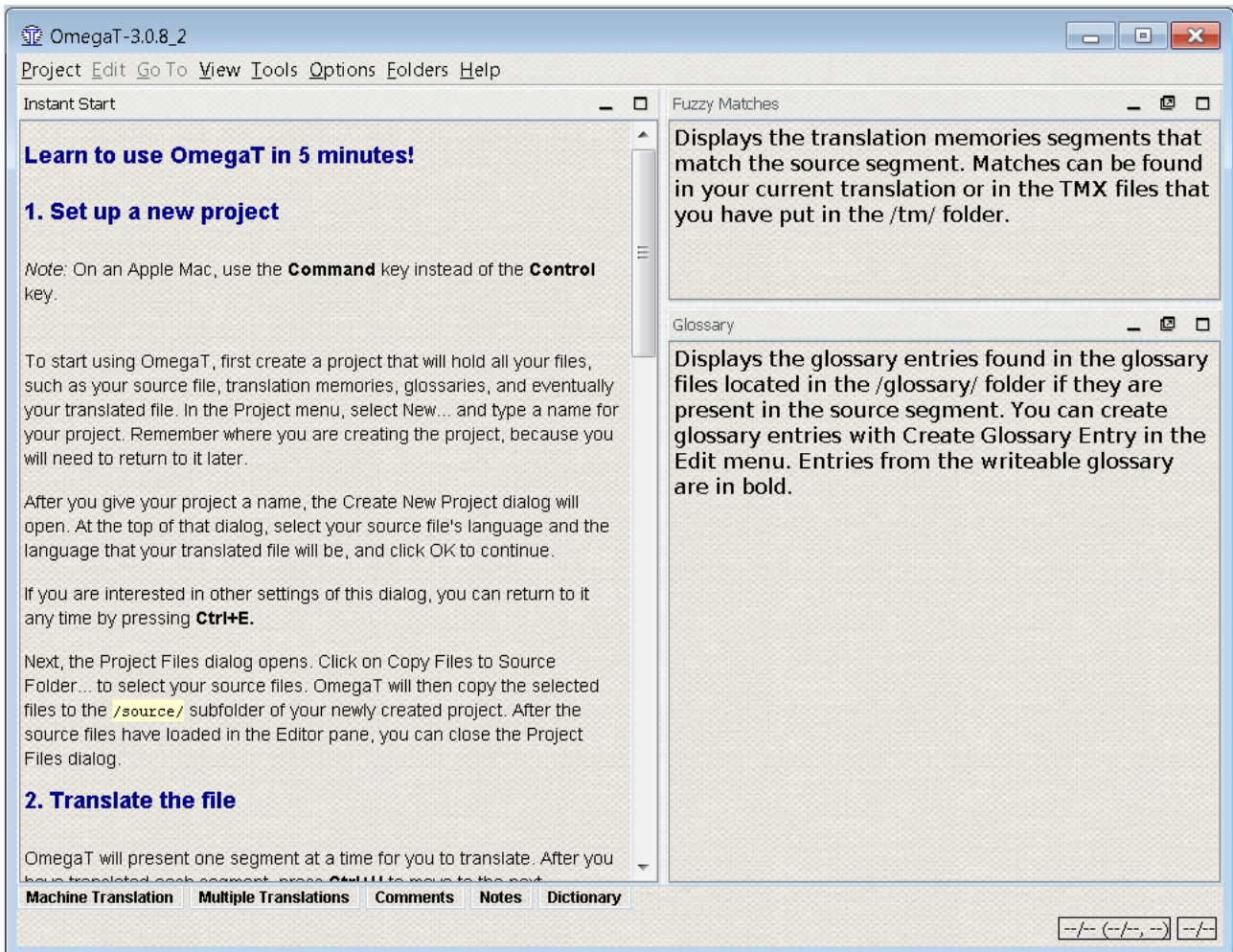
After downloading OmegaT (Windows version), launch the .exe file and follow the on-screen instructions, which should be self-explanatory.

## **3. The OmegaT user interface**

After launching OmegaT, the main OmegaT window should appear with three main panes: the Editing pane, the Fuzzy Matches pane, and the Glossary pane; and five optional panes: Machine Translation, Multiple Translations, Notes, Comments, and Dictionary. The optional panes are covered in the comprehensive User's Manual.

If you can't see all the main panes, drag the borders between them across your screen until they are arranged approximately as shown below.

The Editing pane displays the "Instant Start" tutorial in English. The User's Manual can be called up with the "Help" menu item or F1.



If the Editing pane is empty, the language of your Windows installation is probably not English and OmegaT does not have an Instant Start tutorial in your language. In this case, call up the OmegaT documentation by selecting Help > User's Manual, or by pressing F1.

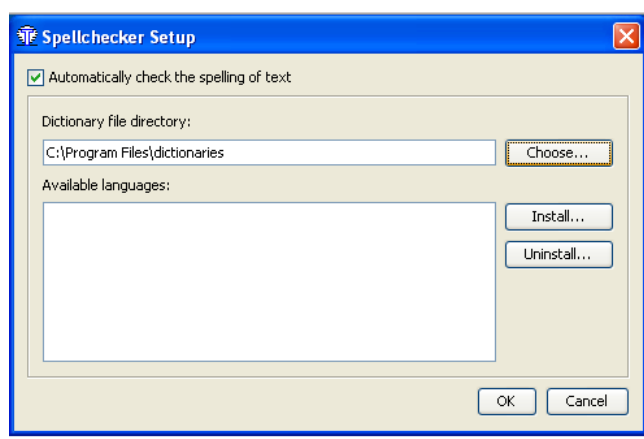
Be warned that the documentation may be a little out of date, owing to the pace at which OmegaT is being developed. For some languages, it is quite a bit out of date. Volunteers to translate the current documentation are welcome! Please contact the Yahoo user group.

## 4. Installing a spelling dictionary

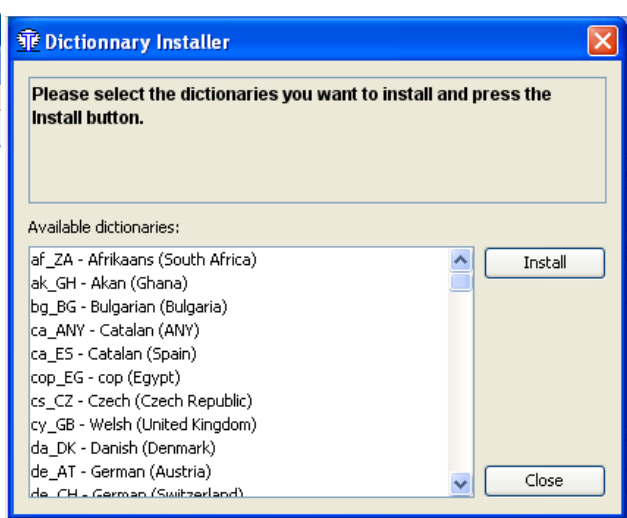
You will need a place to store your dictionary or dictionaries for your target language's spell checker (this is *not* the place for bilingual dictionaries). An example is to create a "dictionaries" folder in your Program Files (Start > My Computer > Local Disk (C:) > Program Files > right click > New Folder, name it "dictionaries.") Then from the OmegaT menu, click Options > Spell checking. Click Choose and navigate to the "dictionaries" folder (screenshot, **a**, below).

Click "Install new dictionary" and you will be given a list of many possible dictionaries (**b**). (If nothing happens, select a different repository – see

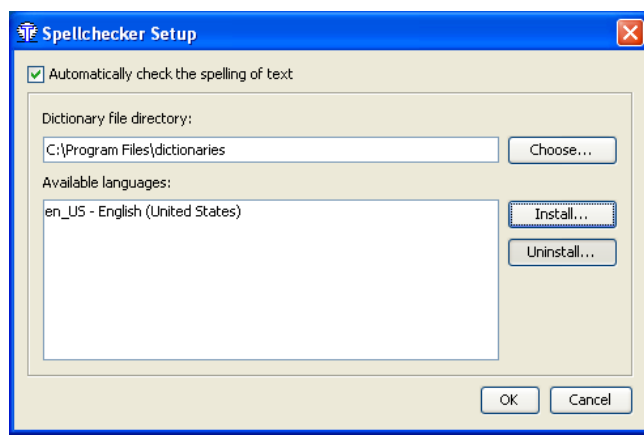
[www.omegat.org/en/howtos/spelling.html](http://www.omegat.org/en/howtos/spelling.html).) Click on the one or ones you want and click Install again, then close the box. Selecting a dictionary and “Install” causes the relevant dictionary to be downloaded from the Internet to the folder which you specified in **a)**. Of course, you must be connected to the Internet for this function to work. You will see that the dictionary or dictionaries appear listed in the “spellchecker setup” dialog window **c)**. Make sure the language variant of your dictionary matches the language variant of your target language, or the spellchecker will not work properly (e.g., British English or American English).



**a)**



**b)**



**c)**

## 5. Creating a project

Most CAT tools use the term “project” to refer to a collection of folders and files associated with a translation. A “project” is usually synonymous with a translation job.

To create a project in OmegaT, select Project, then click New. The “Create a New Project” dialog appears (not shown here).

Navigate to a suitable folder in which to create the project, e.g., the main

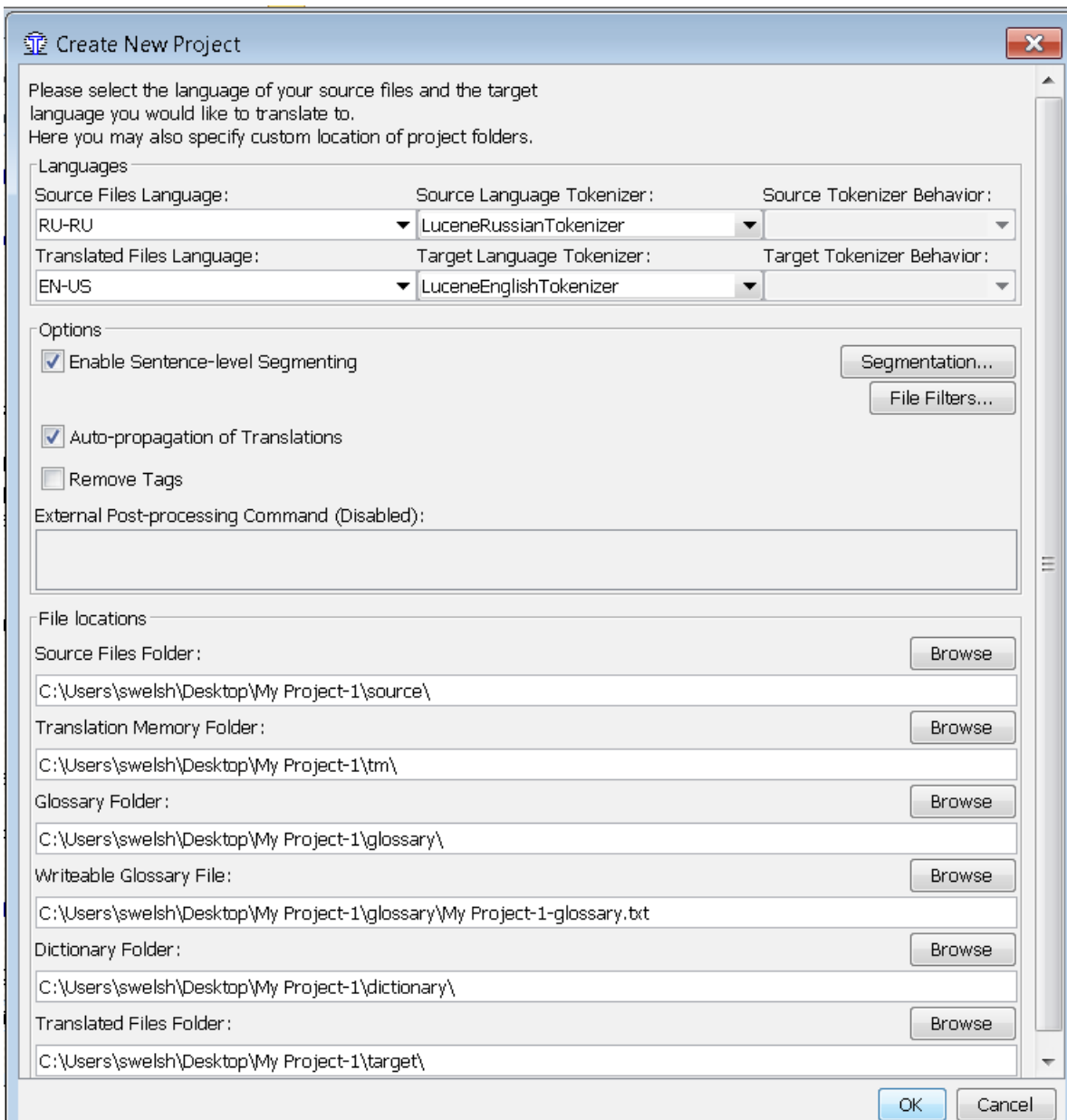


folder you use for your translations, or you can place it on the Desktop for temporary ease of access, and later move it to My Documents or a similar location for storage.

Enter a suitable name for your project, such as the client’s name and purchase order number, and click Save. For purposes of this tutorial, let’s call it **My Project-1**.

The “Create New Project” dialog appears.

Select the desired source and target languages by clicking on the arrows to the right of these fields.



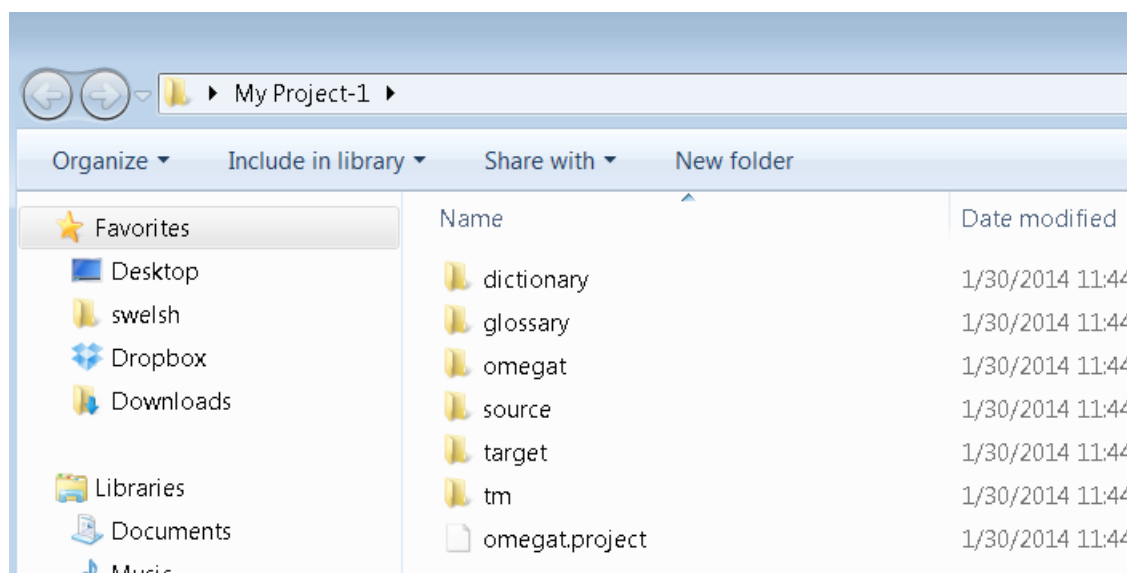
Click on “Segmentation” to call up the Segmentation Setup dialog. OmegaT is supplied with a number of very basic segmentation patterns. Pick the one for the language of your source text (e.g., Italian, “IT.\*”), and use “Move Up” to move it to the top of the list. If the language of your source text isn’t listed, pick a language that is, double-click on the Language Name field, and change it to the abbreviation for your language. Do the same with the Language

Pattern field. If you use OmegaT for any length of time in the future, you will want to customize the segmentation rules. It is not worth getting bogged down with fine-tuning the segmentation rules at this stage; the basic rules as supplied should work adequately for demonstration purposes, even for a different (source) language, provided its sentence punctuation rules are broadly similar (i.e., don't use the Japanese segmentation rules if you are translating from Italian).

Confirm the segmentation rules and then creation of the project with OK. *Do not change any other settings, for the purposes of this tutorial. Do not click on "Browse" or otherwise change the default locations of the project folders (directories).* You have now created an empty project.

The Project Files window is displayed, which is empty since your project does not yet contain any files. (Do not use the "Import Source Files" function during this exercise; you can experiment with it later, on your own.) You can close this window.

Find your project in your file manager: it is a folder with the name of the project, containing the sub-folders */dictionary*, */glossary*, */omegat*, */source*, */target*, */omegat*, and */tm*, and */glossary*. They are all empty. You will put your document to be translated into the */source* folder; the */target* folder will be filled with your translated text; the */omegat* folder need not concern you now; the */tm* and */glossary* files will remain empty unless and until you create a *.tmx* file and a glossary and put them into those respective folders; the */dictionary* folder is where you can put bilingual dictionaries, if you wish, in the appropriate format (see User's Manual)..



## 6. Translating

We provide here two projects for self-training.

The first, to get you started, uses a simple MS Word file in the source language. This does not demonstrate OmegaT's full capabilities, but is intended to help the "newbie" grasp the most essential procedures. The

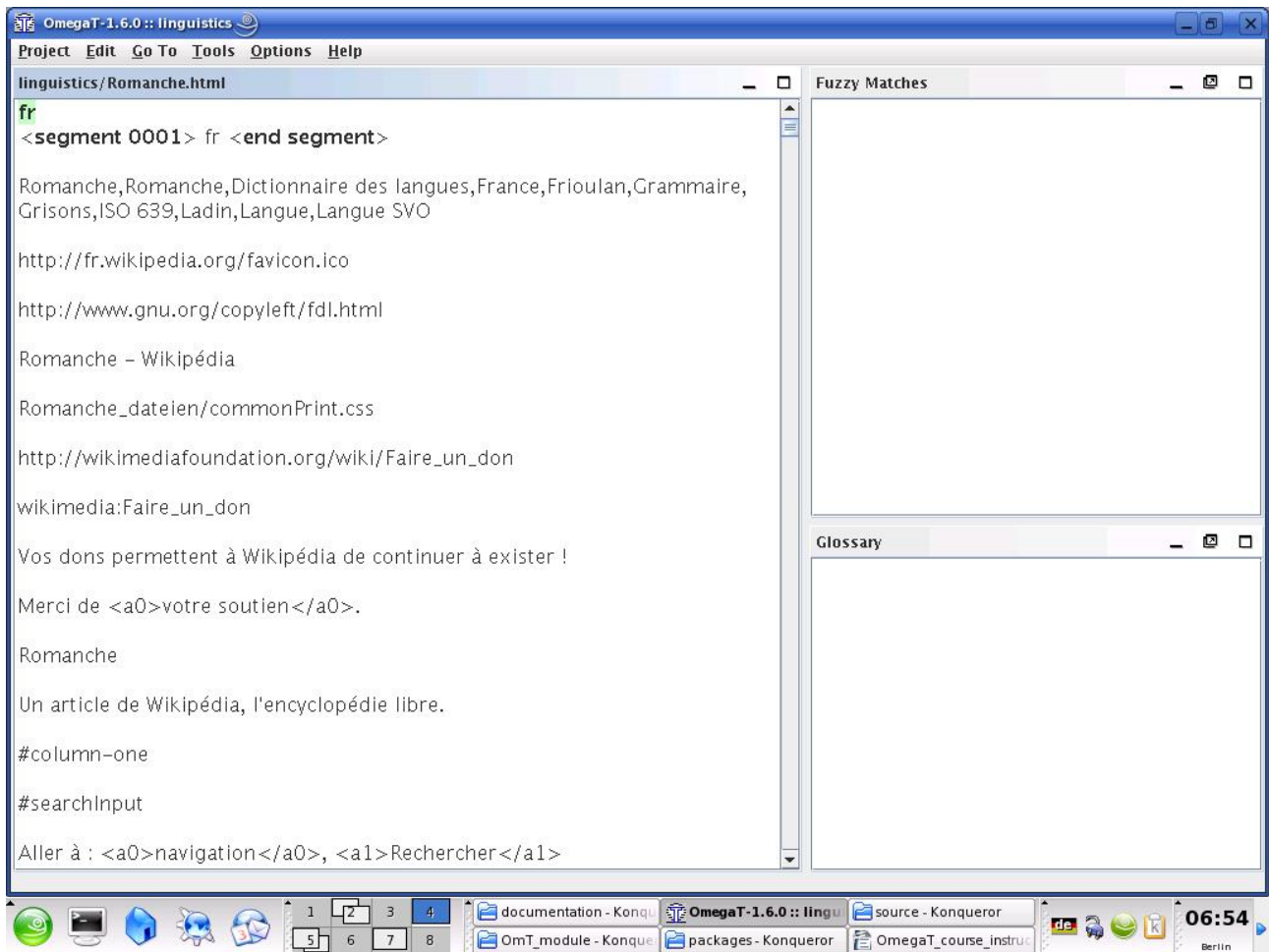
second project uses an HTML file or set of files, downloaded from the Internet, and demonstrates OmegaT's capability to deal with graphics, complex layout, and complex file structure. There you will learn how translation memories, glossaries, and the search function work. These capabilities can be used regardless of the file type (.docx, .xlsx, .odt, .txt, .html, etc.). See **6.B.1-6.B.4**, below. We ignore these in Project 1 only for training purposes. ***Please proceed to Project 2 after completing Project 1!***

*Important:* the purpose of this introduction to OmegaT is to acquaint you with the basic concepts of translation memory software. References to OmegaT's more advanced or obscure functions have deliberately been omitted or kept to a minimum, since a description of them would almost certainly lead to you, as a new user, being unable to see the forest for the trees. Consult the User's Manual for further information.

## **6.A. Project 1: A simple MS Word file**

Launch Open Office Writer and create a new text document. Then use File > Open to Choose open any short MS Word document (.docx format) document in the source language. It should include at least some formatting, such as title, subheads, font changes. (For the purposes of this tutorial, you may use .docx files according to this procedure, although in fact .docx files do not require the mediation of Open Office Writer. See User's Manual when you are ready for the fine points.)File > Save As, name the file, and close it in .odt format. Click and drag the newly created .odt file to the */source* folder of your **My Project-1** folder, which you created in Step 5.

Launch OmegaT and click Project > Open. Navigate to **My Project-1** (it will have the OmegaT icon beside it). Double-click on the file. The Project Files dialog will display your source document. Close the dialog box. Your translation file is now displayed in the Editing pane, and you can begin translating. (Note that this screenshot refers to OmegaT 1.6.0 in the blue bar at the top. Owing to the pace at which OmegaT is being developed, the appearance of the screenshots and possibly some other information may have changed slightly.)



The basic idea is that the text is presented to you one segment at a time. A segment normally corresponds to one sentence. You type your translation into the field provided, confirm by hitting Enter, and so on. OmegaT is an “inline” translation memory application, meaning that you see only one copy of the text; segments which have already been translated appear as such, otherwise the untranslated source text is displayed. Only in the case of the “active” segment, which is highlighted, are both source text and translation displayed.

By default, the source text is pasted into the target text segment for you. This is convenient with heavily tagged (formatted) text, since it is easier to retain the tags undamaged by simply replacing the text between them. With less heavily tagged text, you can delete the content of the target text field and simply type in the translation, or you can instruct OmegaT to leave it blank by default (refer to the User’s Manual for details).

### 6.A.1. Handling tags

Handling tags takes some practice. The general principle is that you translate the text between tags and leave the tags untouched. For example, this text in English:

Look at **that!**

might appear like this in OmegaT:

Look at `<a0>that</a0>!`

and be translated like this in OmegaT into German:

Schau dir `<a0>das</a0>` an!

which then appears like this in the final text:

Schau dir **das** an!

In this case, `<a0>` and `</a0>` are the opening and closing tags for bold text respectively. “`<a0>`” and “`</a0>`” do not necessarily indicate the beginning and end of bold text, however; if the word “that” were in italics in the source text rather than bold, the tags would be the same. Only by looking at the source text can you ascertain the actual function of the tags in a particular case.

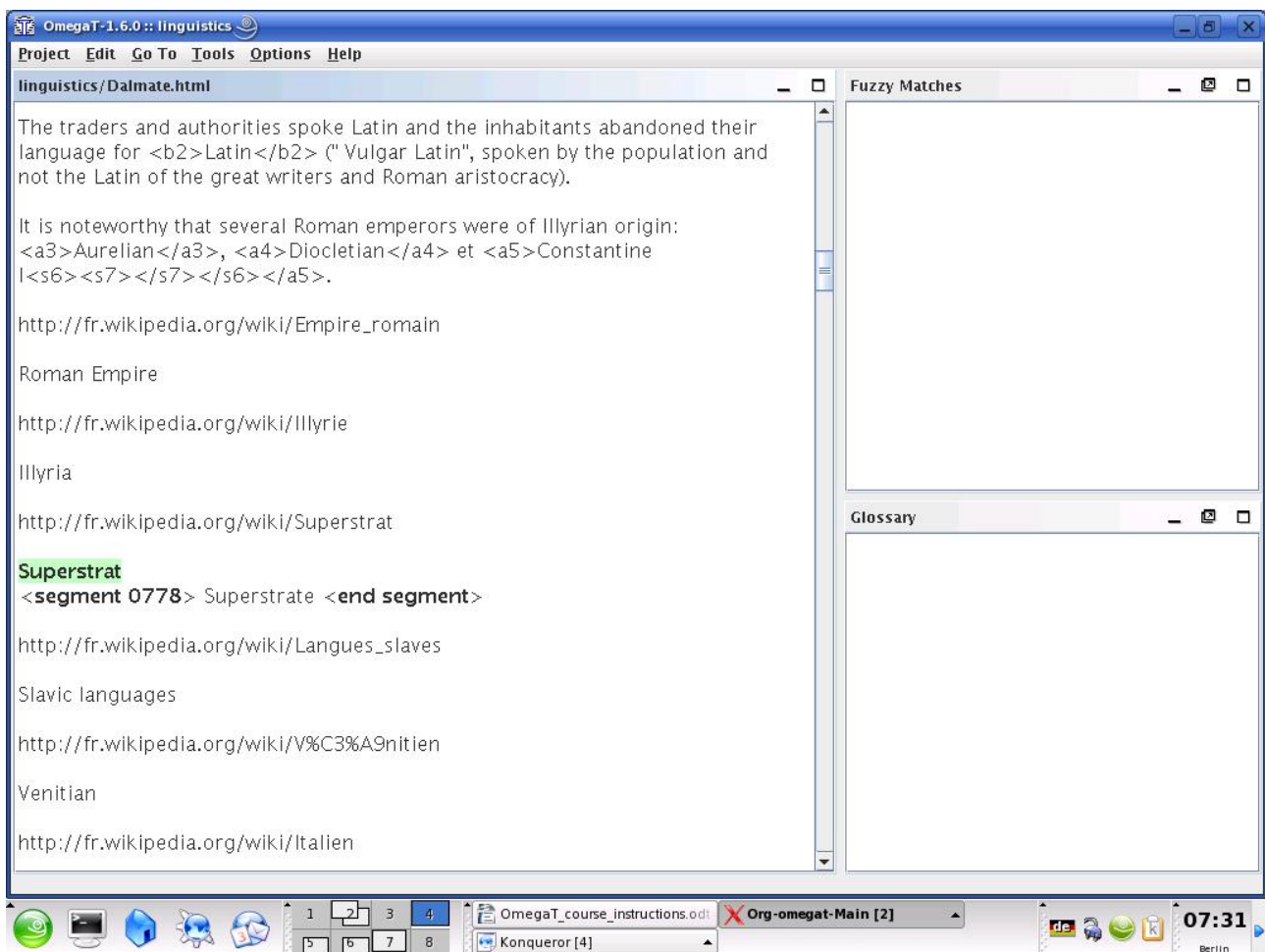
The current Microsoft Office file formats (.docx, .xlsx, .pptx) can generate a large number of unnecessary tags, making translation cumbersome. One solution to this problem is Codezapper (<http://asap-traduction.com/CodeZapper>), which can eliminate these, while keeping the tags you actually need; or you can use the Remove Tags feature of OmegaT itself to remove *all* tags. See the User’s Manual (F1).

Until you are familiar with tag handling, it is best to retain all the tags and to keep them in the same order if at all possible. OmegaT does allow you to delete tags and to change their order, but only if certain rules are carefully followed. Otherwise, you may risk corrupting your final document and preventing it from being opened. For details of the tag handling rules, refer to the User’s Manual.

## 6.A.2. Checking your translation

Eventually, you will reach the end of the file you are translating. If your project contains more than one file for translation, confirming the last segment in the first file will take you to the topmost segment in the second file.

An advantage probably common to all translation memory applications is that they simplify checking: once you have completed the first draft of your translation, you can “walk” through it with the source and target versions of each sentence displayed next to each other. In the case of OmegaT, the source is displayed above the target text of the active segment, facilitating comparison:



### 6.A.3. Validating tags, creating target documents

After you have checked and saved in your translation, newer versions of OmegaT will check automatically that you have not damaged any tags during the translation process, or you can do it manually yourself: Tools > Validate Tags. *Note: Damaged or missing tags in XML file formats, such as OpenOffice.org, are likely to result in the output file failing to open at all.*

Create the target documents by selecting Project > Create Translated Documents. The translated documents are automatically generated in their original format in the */target* folder. If your project consists of multiple source files, possibly in multiple sub-folders and with accompanying graphics files, etc., as is the case with the examples in our second, HTML, project, the entire structure of the files in */source* is reproduced in */target*.

### 6.A.4. Final checking, correction, and delivery

If you prefer to check your translation on paper, print out the target document. *Don't, however, make corrections in the Open Office Writer or MS Word file: find the relevant segments in OmegaT (using the text search function, to be discussed below, Step 6.B.4) and make them there.*

After revising the text and making any corrections, create the translated

documents again and close the project.

You are now ready to deliver the job. If the client requires MS Office format, simply save the target file in OpenOffice.org in the relevant MS format (.doc, etc.)

## **6.B. Project 2: An HTML document**

### **6.B.1. Downloading the supplementary materials**

Launch OmegaT and create a new project. Let's call it **My Project-2**. Set it up just as you did with **My Project-1**. The procedures described in Project 1, Steps **6.A.1-6.A.4**, all apply here too, for HTML files.

All files selected for translation in this tutorial are from Wikipedia; some projects comprise more than one file, to familiarize you with OmegaT's usefulness for such structured documents.

The table provides a summary of web pages available, for which our tutorial's "legacy" files have been constructed:

<b>Language combination</b>	<b>Subject</b>	<b>Source</b>
English to Chinese	stocks	<a href="https://en.wikipedia.org/wiki/Share_%28finance%29">en.wikipedia.org/wiki/Share_%28finance%29</a> <a href="https://en.wikipedia.org/wiki/Shareholder">en.wikipedia.org/wiki/Shareholder</a>
English to Czech	biocoenosis	<a href="https://en.wikipedia.org/wiki/Biocoenosis">en.wikipedia.org/wiki/Biocoenosis</a> <a href="https://en.wikipedia.org/wiki/Phytosociology">en.wikipedia.org/wiki/Phytosociology</a>
English to Dutch	stocks	<a href="https://en.wikipedia.org/wiki/Share_%28finance%29">en.wikipedia.org/wiki/Share_%28finance%29</a> <a href="https://en.wikipedia.org/wiki/Shareholder">en.wikipedia.org/wiki/Shareholder</a>
English to Dutch	russian-miscellaneous	<a href="https://en.wikipedia.org/wiki/Russian">en.wikipedia.org/wiki/Russian</a> <a href="https://en.wikipedia.org/wiki/Russian_American">en.wikipedia.org/wiki/Russian_American</a> <a href="https://en.wikipedia.org/wiki/Russian_Canadian">en.wikipedia.org/wiki/Russian_Canadian</a> <a href="https://en.wikipedia.org/wiki/Rusky_Island">en.wikipedia.org/wiki/Rusky_Island</a>
English to French	swimming	<a href="https://en.wikipedia.org/wiki/Individual_medley">en.wikipedia.org/wiki/Individual_medley</a> <a href="https://en.wikipedia.org/wiki/World_records_in_swimming">en.wikipedia.org/wiki/World_records_in_swimming</a>
English to French	dorset	<a href="https://en.wikipedia.org/wiki/Jurassic_Coast">en.wikipedia.org/wiki/Jurassic_Coast</a> <a href="https://en.wikipedia.org/wiki/Old_Harry_Rocks">en.wikipedia.org/wiki/Old_Harry_Rocks</a> <a href="https://en.wikipedia.org/wiki/Durdle_Door">en.wikipedia.org/wiki/Durdle_Door</a>
English to French	medicine	<a href="https://en.wikipedia.org/wiki/Hippocrates">en.wikipedia.org/wiki/Hippocrates</a>
English to German	hilton	<a href="https://en.wikipedia.org/wiki/Hilton_Hotels">en.wikipedia.org/wiki/Hilton_Hotels</a> <a href="https://en.wikipedia.org/wiki/Great_Western_Hotel,_London">en.wikipedia.org/wiki/Great_Western_Hotel,_London</a> <a href="https://en.wikipedia.org/wiki/Waldorf_Hilton">en.wikipedia.org/wiki/Waldorf_Hilton</a> <a href="https://en.wikipedia.org/wiki/The_London_Hilton_on_Park_Lane">en.wikipedia.org/wiki/The_London_Hilton_on_Park_Lane</a>
English to German	construction	<a href="https://en.wikipedia.org/wiki/Wall">en.wikipedia.org/wiki/Wall</a> <a href="https://en.wikipedia.org/wiki/Panelling">en.wikipedia.org/wiki/Panelling</a>
English to Italian	russian-miscellaneous	<a href="https://en.wikipedia.org/wiki/Russian">en.wikipedia.org/wiki/Russian</a> <a href="https://en.wikipedia.org/wiki/Russian_American">en.wikipedia.org/wiki/Russian_American</a> <a href="https://en.wikipedia.org/wiki/Russian_Canadian">en.wikipedia.org/wiki/Russian_Canadian</a> <a href="https://en.wikipedia.org/wiki/Rusky_Island">en.wikipedia.org/wiki/Rusky_Island</a>
English to Italian	yoga	<a href="https://en.wikipedia.org/wiki/Karma_yoga">en.wikipedia.org/wiki/Karma_yoga</a> <a href="https://en.wikipedia.org/wiki/Jnana_yoga">en.wikipedia.org/wiki/Jnana_yoga</a>
English to Italian	civil-engineering	<a href="https://en.wikipedia.org/wiki/Blind_Jack">en.wikipedia.org/wiki/Blind_Jack</a>
English to Polish	construction	<a href="https://en.wikipedia.org/wiki/Wall">en.wikipedia.org/wiki/Wall</a> <a href="https://en.wikipedia.org/wiki/Panelling">en.wikipedia.org/wiki/Panelling</a>
English to Polish	poland	<a href="https://en.wikipedia.org/wiki/Lubusz_Voivodship">en.wikipedia.org/wiki/Lubusz_Voivodship</a> <a href="https://en.wikipedia.org/wiki/Podlasie_Voivodship">en.wikipedia.org/wiki/Podlasie_Voivodship</a>
English to Polish	medicine	<a href="https://en.wikipedia.org/wiki/Hippocrates">en.wikipedia.org/wiki/Hippocrates</a>
English to Polish	swimming	<a href="https://en.wikipedia.org/wiki/Individual_medley">en.wikipedia.org/wiki/Individual_medley</a> <a href="https://en.wikipedia.org/wiki/World_records_in_swimming">en.wikipedia.org/wiki/World_records_in_swimming</a>
English to Portuguese	russian-miscellaneous	<a href="https://en.wikipedia.org/wiki/Russian">en.wikipedia.org/wiki/Russian</a> <a href="https://en.wikipedia.org/wiki/Russian_American">en.wikipedia.org/wiki/Russian_American</a> <a href="https://en.wikipedia.org/wiki/Russian_Canadian">en.wikipedia.org/wiki/Russian_Canadian</a> <a href="https://en.wikipedia.org/wiki/Rusky_Island">en.wikipedia.org/wiki/Rusky_Island</a>
English to Russian	biocoenosis	<a href="https://en.wikipedia.org/wiki/Biocoenosis">en.wikipedia.org/wiki/Biocoenosis</a> <a href="https://en.wikipedia.org/wiki/Phytosociology">en.wikipedia.org/wiki/Phytosociology</a>



English to Russian	stocks	<a href="https://en.wikipedia.org/wiki/Share_%28finance%29">en.wikipedia.org/wiki/Share_%28finance%29</a> <a href="https://en.wikipedia.org/wiki/Shareholder">en.wikipedia.org/wiki/Shareholder</a>
English to Spanish	dorset	<a href="https://en.wikipedia.org/wiki/Jurassic_Coast">en.wikipedia.org/wiki/Jurassic_Coast</a> <a href="https://en.wikipedia.org/wiki/Old_Harry_Rocks">en.wikipedia.org/wiki/Old_Harry_Rocks</a> <a href="https://en.wikipedia.org/wiki/Durdle_Door">en.wikipedia.org/wiki/Durdle_Door</a>
English to Spanish	appliances	<a href="https://en.wikipedia.org/wiki/Small_appliance">en.wikipedia.org/wiki/Small_appliance</a> <a href="https://en.wikipedia.org/wiki/Kitchen_appliance">en.wikipedia.org/wiki/Kitchen_appliance</a> <a href="https://en.wikipedia.org/wiki/Major_appliance">en.wikipedia.org/wiki/Major_appliance</a>
English to Spanish	cereal	<a href="https://en.wikipedia.org/wiki/Rye">en.wikipedia.org/wiki/Rye</a> <a href="https://en.wikipedia.org/wiki/Sorghum">en.wikipedia.org/wiki/Sorghum</a>
English to Spanish	skye	<a href="https://en.wikipedia.org/wiki/Skye">en.wikipedia.org/wiki/Skye</a>
English to Spanish	hilton	<a href="https://en.wikipedia.org/wiki/Hilton_Hotels">en.wikipedia.org/wiki/Hilton_Hotels</a> <a href="https://en.wikipedia.org/wiki/Great_Western_Hotel,_London">en.wikipedia.org/wiki/Great_Western_Hotel,_London</a> <a href="https://en.wikipedia.org/wiki/Waldorf_Hilton">en.wikipedia.org/wiki/Waldorf_Hilton</a> <a href="https://en.wikipedia.org/wiki/The_London_Hilton_on_Park_Lane">en.wikipedia.org/wiki/The_London_Hilton_on_Park_Lane</a>
English to Turkish	hilton	<a href="https://en.wikipedia.org/wiki/Hilton_Hotels">en.wikipedia.org/wiki/Hilton_Hotels</a> <a href="https://en.wikipedia.org/wiki/Great_Western_Hotel,_London">en.wikipedia.org/wiki/Great_Western_Hotel,_London</a> <a href="https://en.wikipedia.org/wiki/Waldorf_Hilton">en.wikipedia.org/wiki/Waldorf_Hilton</a> <a href="https://en.wikipedia.org/wiki/The_London_Hilton_on_Park_Lane">en.wikipedia.org/wiki/The_London_Hilton_on_Park_Lane</a>
French to English	seine-et-marne	<a href="https://fr.wikipedia.org/wiki/D%C3%A9mographie_de_Seine-et-Marne">fr.wikipedia.org/wiki/D%C3%A9mographie_de_Seine-et-Marne</a> <a href="https://fr.wikipedia.org/wiki/Seine-et-Marne">fr.wikipedia.org/wiki/Seine-et-Marne</a>
French to English	esoteric	<a href="https://fr.wikipedia.org/wiki/Augure">fr.wikipedia.org/wiki/Augure</a> <a href="https://fr.wikipedia.org/wiki/Divination">fr.wikipedia.org/wiki/Divination</a> <a href="https://fr.wikipedia.org/wiki/Effet_Barnum">fr.wikipedia.org/wiki/Effet_Barnum</a>
French to English	dancing	<a href="https://fr.wikipedia.org/wiki/Danse">fr.wikipedia.org/wiki/Danse</a>
French to English	cardiology	<a href="https://fr.wikipedia.org/wiki/Cardiologie">fr.wikipedia.org/wiki/Cardiologie</a> <a href="https://fr.wikipedia.org/wiki/Tilt-test">fr.wikipedia.org/wiki/Tilt-test</a>
French to English	linguistics	<a href="https://fr.wikipedia.org/wiki/Romanche">fr.wikipedia.org/wiki/Romanche</a> <a href="https://fr.wikipedia.org/wiki/Dalmate">fr.wikipedia.org/wiki/Dalmate</a>
French to English	ornithology	<a href="https://fr.wikipedia.org/wiki/Moineau">fr.wikipedia.org/wiki/Moineau</a> <a href="https://fr.wikipedia.org/wiki/Moineau_domestique">fr.wikipedia.org/wiki/Moineau_domestique</a>
French to English	wine	<a href="https://fr.wikipedia.org/wiki/Vin_blanc">fr.wikipedia.org/wiki/Vin_blanc</a> <a href="https://fr.wikipedia.org/wiki/Vin_rouge">fr.wikipedia.org/wiki/Vin_rouge</a>
French to English	rugby	<a href="https://fr.wikipedia.org/wiki/Rugby">fr.wikipedia.org/wiki/Rugby</a> <a href="https://fr.wikipedia.org/wiki/William_Webb_Ellis">fr.wikipedia.org/wiki/William_Webb_Ellis</a>
German to English	canterbury	<a href="https://de.wikipedia.org/wiki/Canterbury">de.wikipedia.org/wiki/Canterbury</a> <a href="https://de.wikipedia.org/wiki/Canterbury_(Begriffserkl%C3%A4rung)">de.wikipedia.org/wiki/Canterbury_(Begriffserkl%C3%A4rung)</a>
German to English	domestication	<a href="https://de.wikipedia.org/wiki/Domestizierung">de.wikipedia.org/wiki/Domestizierung</a>
German to English	equestrianism	<a href="https://de.wikipedia.org/wiki/Dressurreiten">de.wikipedia.org/wiki/Dressurreiten</a> <a href="https://de.wikipedia.org/wiki/Hohe_Schule_(Reitsport)">de.wikipedia.org/wiki/Hohe_Schule_(Reitsport)</a>
German to English	railways	<a href="https://de.wikipedia.org/wiki/Neigetechnik">de.wikipedia.org/wiki/Neigetechnik</a>
German to English	music	<a href="https://de.wikipedia.org/wiki/Oberton">de.wikipedia.org/wiki/Oberton</a>

German to English	psychiatry	de.wikipedia.org/wiki/Psychiatrie
German to English	humour	de.wikipedia.org/wiki/Fawlty_Towers
German to English	teaching	de.wikipedia.org/wiki/Lehrer
Italian to English	librarianship	it.wikipedia.org/wiki/Biblioteca
Italian to English	politics	it.wikipedia.org/wiki/Gianni_De_Michelis it.wikipedia.org/wiki/Mariano_Rumor
Portuguese to English	chagas	pt.wikipedia.org/wiki/Doen%C3%A7a_de_Chagas
Russian to English	law	ru.wikipedia.org/wiki/%D0%9F %D1%80%D0%B0%D0%B2%D0%BE
Spanish to English	montevideo	es.wikipedia.org/wiki/Montevideo es.wikipedia.org/wiki/Bruno_Mauricio_de_Zabala
Spanish to English	unicef	es.wikipedia.org/wiki/Fondo_de_Naciones_Unidas_ para_la_Infancia es.wikipedia.org/wiki/Niño es.wikipedia.org/wiki/Derechos_del_niño
Spanish to English	climbing	es.wikipedia.org/wiki/Escalada_en_hielo

Once you have decided on a subject, download the files from the Internet, navigating so that they go directly into the */source* folder of **My Project-2**. Make sure you download the entire web page, including graphics. (Various browsers have slightly different methods for doing this. For Internet Explorer 7.x, for example, File > Save As> Webpage, Complete; for Firefox 5, File > Save Page as > Web Page, Complete.)

Your */source* folder will now contain one or more HTML file(s), together with associated files such as graphics files. The structure of folders and subfolders must be retained. You should be able to open the HTML files in your browser and view them approximately as they appear on the Internet.

Next, download the “legacy package” which we have created for use in this tutorial, which is at:

[www.omegat.org/training/materials/legacy.zip](http://www.omegat.org/training/materials/legacy.zip)

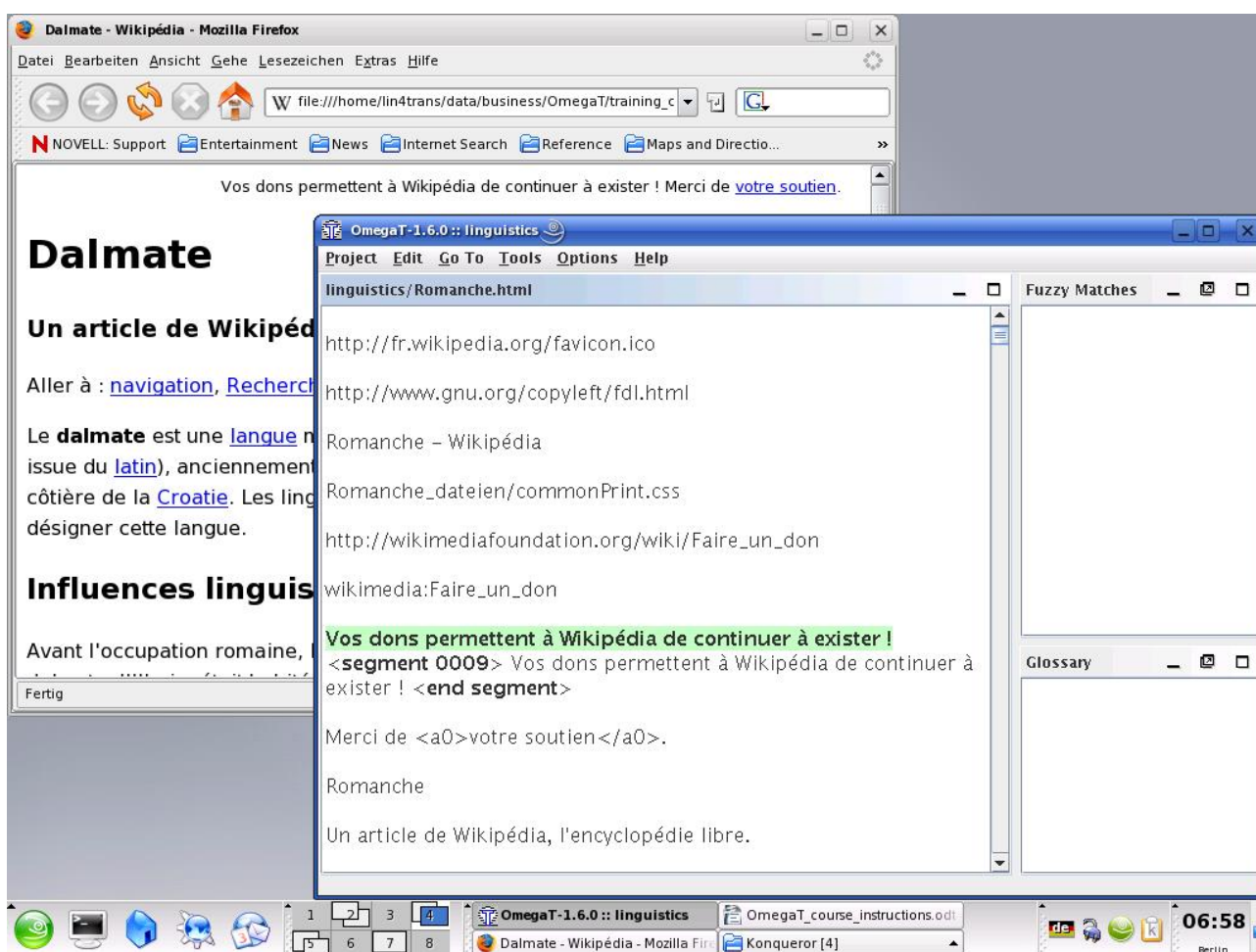
Unpack the legacy.zip archive to any suitable temporary location (such as My Documents or the Desktop). It contains a number of translation memory files (with the extension .tmx) and OmegaT glossary files (with the extension .txt). Identify the .tmx file and the glossary (.txt) file for your language combination. Copy the .tmx file to the */tm* folder and the .txt file to the */glossary* folder of your OmegaT project.

***Some more points to note about the HTML files for translation:***

- If you view these files in your browser on your PC, they will differ slightly in appearance from the pages on the Internet. This has nothing to do with OmegaT or your browser, but is due to the fact that Wikipedia does not allow all the template information for its pages to be downloaded.
- Wikipedia is subject to continual change. The pages listed here have been selected as suitable examples (and the “legacy” files have been created to correspond to them), but changes to them may make them

less suitable, and they may even disappear or move elsewhere.

- When you are translating from HTML, you will see a lot of codes that do not need to be translated, particularly at the beginning of each file. In many cases, whole segments (“sentences”), such as web links, can be left as they are. Simply click through these by pressing Enter to confirm that the target is to remain the same.
- HTML is generally very tolerant of corrupted or missing tags; the result is likely to be a file which will open, but with unexpected or undesired formatting results. Review Step **6.A.1** on handling tags.



- With heavily tagged HTML text, you may find it convenient to keep the source text open at the same time and to toggle between the two (Alt+tab) in order to be able to see more quickly what text actually has to be translated. As a general rule for the purpose of this tutorial, assume that if something is visible in your browser, it should be translated.

The windows in the above screenshot have been reduced in size in order to show the idea of keeping OmegaT and the source text (in your browser) open at the same time.

- When you are finished and ready to check your work, you can open the

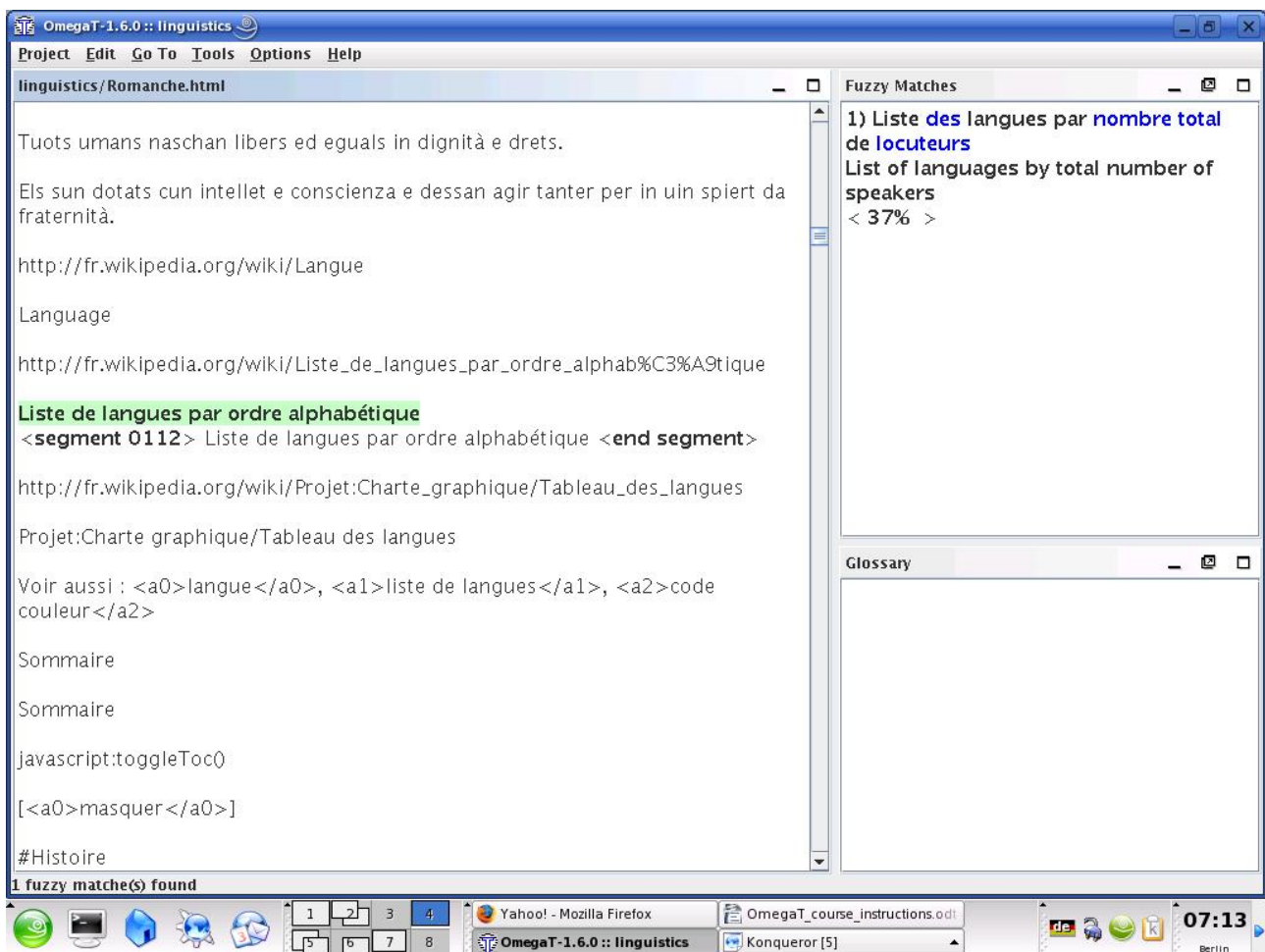
translated files in */target* in your browser and read through them, and/or print them out for checking. If you prefer to read through the finished text on the screen, i.e., in your browser, toggle between the browser and OmegaT. When you find errors in the text in the browser window, make the corresponding changes within OmegaT.

## **6.B.2. Translation memory (fuzzy match function)**

Eventually, you will encounter fuzzy matches. Normally, it may be some time before you encounter useful fuzzy matches, but the “legacy” translation memory file has been provided for you in this tutorial to ensure that you will find at least some.

“Fuzzy matches” are segments with some degree of similarity to the segment you are translating. They may be within the current project (either in the text you are currently translating or, where a project comprises more than one source text, in another text within the same project), or they may be in the legacy .tmx file(s) which you placed in the */tm* folder. Since the legacy .tmx files in the */tm* folder are the product of previous translation jobs, you will not normally have any useful legacy .tmx files to call upon until you have completed a few jobs in OmegaT, and placed them into the */tm* folder of your new job.

Here, the current segment shows a fuzzy match with a segment elsewhere in the project that has already been translated:



If after looking at the fuzzy match you decide that it is useful, you can paste it into the active segment at any point (Ctrl+I), or replace (overwrite) the active segment with it completely (Ctrl+R) and then modify it, or copy and paste parts of it into the active segment (Ctrl+C and Ctrl+V to copy and paste).

The legacy material provided for the purpose of this tutorial is intended to demonstrate the functions of OmegaT. Under no circumstances should the translations in the translation memory be assumed to be reliable.

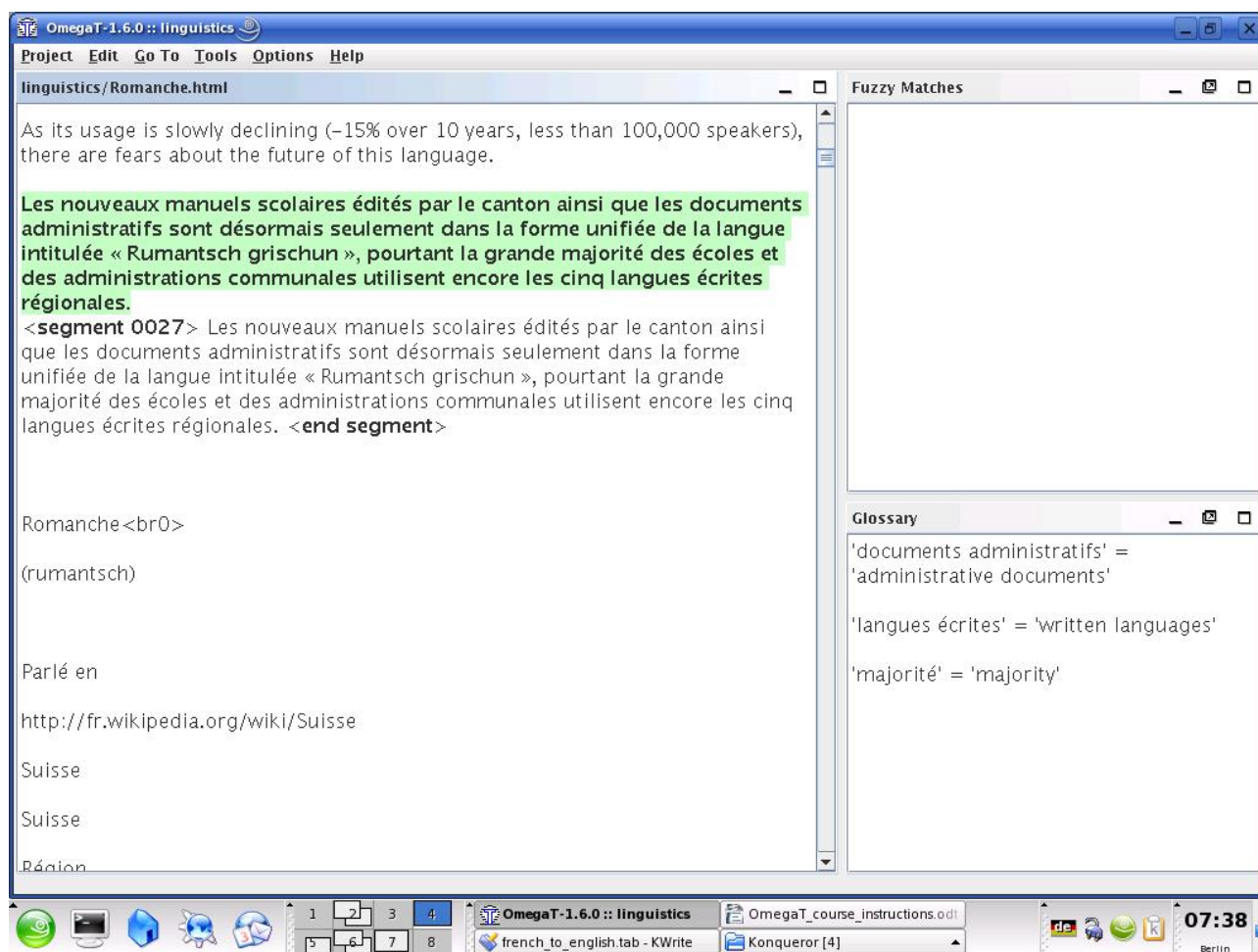
With heavily tagged text, you are likely to find that OmegaT yields numerous spurious fuzzy matches in which only the tags are identical, and which are not therefore useful.

### 6.B.3. The glossary function

Before beginning translating, you placed a glossary file in the */glossary* folder. Glossary files are simply plain-text files containing two or three columns of terms delimited by tabs. (See User's Manual for how to set up a glossary.)

OmegaT creates an empty default glossary file for a project when the project is first created. You can also add further glossary files of your own. For the purpose of this project, a glossary (.txt) file has been created for you for each language combination. This file will be of limited practical help to you during translation, but it serves to show you how the glossary terms are displayed to

you in OmegaT when they occur: immediately in the OmegaT glossary pane. Terms and their translations will appear when you reach a segment containing them.



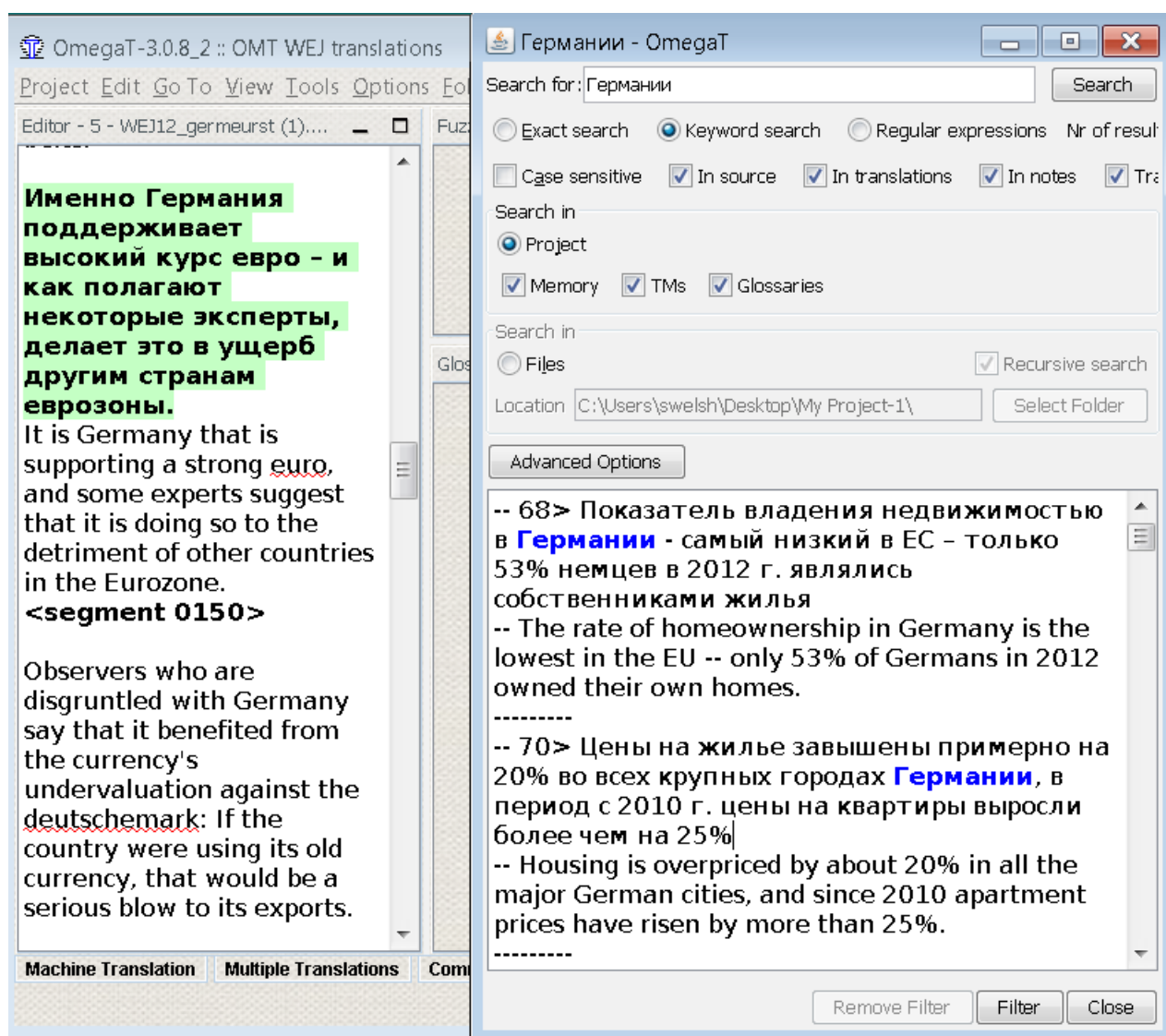
Wild card characters are not accepted. Do not put an article before the noun in a glossary; if you want to indicate the gender of the noun, put that after the first tab in the glossary, in the column where you put the translation. Thus, if your source document contains the words *ein Mensch*, a glossary that has the source text as *der Mensch* will not find it; use *Mensch* for the glossary instead. For more on glossaries, including adding new terms, consult the User's Manual.

For more advanced users, there is a plugin called OmegaT-tokenizers (for use in OmegaT version 2.1.1 and later versions), which allows OmegaT to recognize inflected forms of a word. Those interested may consult the User's Manual and the users' group on Yahoo. The more recent versions of OmegaT automatically incorporate a "tokenizer" function, which allows the system to recognize inflected forms of a word.

## 6.B.4. The text search (find) function

Ctrl+F calls up a search dialog with which you can search for any term

throughout the project, i.e., within the source text and the legacy translation memory files, source segments, target segments, and glossary files.



As with the glossary function, you can use keyboard shortcuts to copy and paste content from the Text Search window into the active translation segment. If you wish to change the text in a segment within the current project which is displayed in the Text Search window, double-click on it and OmegaT will take you to the segment concerned.

The Text Search window has quite powerful search functions. For details, see the User's Manual. Note that there is no an automatic "global replace" function available as a "scripting" plugin outside the main program, also described in the User's Manual., but there are "workarounds" that can be used.

## 7. The OmegaT user group

OmegaT has a friendly and helpful user group. Once you have joined, you can post questions there regarding OmegaT.

The group can be found at [groups.yahoo.com/group/omegat](http://groups.yahoo.com/group/omegat).

The group is hosted by Yahoo! OmT has moderated membership, i.e., your membership first has to be approved. This is in order to prevent spammers accessing the group. To join, simply “Sign In” using your e-mail. You will be required to send a short message from which the moderator can tell that you are a genuine user and not a spammer.

## **8. Appendix**

### **GNU Free Documentation License**

#### **0. Preamble**

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft,” which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

#### **1. APPLICABILITY AND DEFINITIONS**

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document,” below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “you.” You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall



subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque.”

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements,” “Dedications,” “Endorsements,” or “History.”) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ”

according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## **2. VERBATIM COPYING**

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

## **3. COPYING IN QUANTITY**

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document

well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

## **4. MODIFICATIONS**

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- **A.** Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- **B.** List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- **C.** State on the Title page the name of the publisher of the Modified Version, as the publisher.
- **D.** Preserve all the copyright notices of the Document.
- **E.** Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- **F.** Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- **G.** Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- **H.** Include an unaltered copy of this License.
- **I.** Preserve the section Entitled "History," Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- **J.** Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- **K.** For any section Entitled "Acknowledgements" or "Dedications," Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

- **L.** Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- **M.** Delete any section Entitled “Endorsements.” Such a section may not be included in the Modified Version.
- **N.** Do not retitling any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.
- **O.** Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements,” provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

## **5. COMBINING DOCUMENTS**

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise

combine any sections Entitled “Acknowledgements,” and any sections Entitled “Dedications.” You must delete all sections Entitled “Endorsements.”

## **6. COLLECTIONS OF DOCUMENTS**

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## **7. AGGREGATION WITH INDEPENDENT WORKS**

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## **8. TRANSLATION**

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements,” “Dedications,” or “History,” the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## **9. TERMINATION**

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

## **10. FUTURE REVISIONS OF THIS LICENSE**

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.