Sample of a research and recommendation document

The “Localizing Documentaton” document was written for a department of New York Life considering localizing and translating their documentation. I did the research for them about what tools were best and advised them on concepts, terms, and guidelines regarding the localization process.

Help Tool Recommendations and Tips

Localizing Documentation

Lori Kaufman, Senior Technical Writer
New York Life - Agency Department

Contents

Localization Guidelines

Before you consider which help authoring tool to use, there are some issues to consider when beginning a documentation project that will need to be internationalized and localized.

This section contains some tips for localizing documentation obtained from "The Guide to Translation and Localization," by Lingo Systems.

Terms

A help system or user's guide needs to be internationalized first before localization is considered. **Internationalization** is the process of designing and writing a document so that it can be easily and efficiently localized. For example, when laying out a document, leave plenty of white space so that text expansion can be easily accommodated.

Localization is the process of customizing a document for consumers in a target market so that when they use it, they form the impression that it was designed by a native of their own country. This is an important process so that the **Translation** process can be completed smoothly and accurately. Localization must include the development of:

- A style guideline,
- A glossary in the source language, and
- A terminology list in the target language.

Style Guidelines

Style guidelines are writing guidelines that the linguist can follow during the translation process. They are either provided by the author of the English document (client) or are developed by the linguist in coordination with the client. The following are some issues that can be addressed in style guidelines:

- Tone of the localized documentation,
- Those terms that are translated, and those that are not,
- Rules for capitalization and accent marks,
- Translation of titles and subtitles

- Conversion of measurements
- Rules for spelling numbers,
- Use of abbreviations, and
- Punctuation rules.

NOTE: The following is a recommendation directly from the book that I feel is important when considering localizing documentation: "The quality of the localized documentation is largely dependent upon the quality of the source text. Therefore, it is important that the technical writer knows in advance that the documentation is to be translated. It can be beneficial to have the technical writer initially work with the localization vendor to ensure that the subsequent documentation is created for a global audience, that cultural and other country-specific issues are addressed early in the process, and that specific steps are taken to ease the translation process."

Glossary

A **glossary** is a list of words in the source language in which difficult or technical, product-specific terms are explained. Typically, the glossary is developed by the technical writer who is writing the documentation.

Terminology List

A **terminology list** is an agreed-upon list of terms, in the target language, to be used in the localization process. It ensures:

- Consistency of linguistic processes throughout the process (i.e., that all people on the project work with the same terms), covering product, industry, and user interface terms, for use in all of the components,
- Consistency of abbreviations, product names, nontranslated terms, and measurements,
- Consistency between country and company standards,
- Local suitability, and
- Consensus between the client and the localization provider and any other involved parties.

The terminology is based on:

- The product-specific glossary developed by the technical writer of the source document,
- The localized user interface terminology of major software developers,
- Prior localized software and documentation, and
- All other localized resource materials such as marketing collateral, product lists, as well as company and country standards.

Standards

Company "standards" are ways of referring to specific things or specific documents that are unique to the particular company and include part numbers, technical and product support information, warranties, license agreements, copy rights, references to other software, product names, brand names, and nonlocalized components.

Country "standards" are ways of expressing functional or cultural dictates, such as publishing standards, sorting of lists, abbreviations, time, dates, holidays, currency, and measurements.

NOTE: In order to match the company and country standards, and to make sure that there is consistency and accuracy between software and documentation, the terminology list must be developed early on in the localization process before the actual translation begins.

Developing the terminology list

The lead linguist on the project develops the terminology list. He/she gathers all resource materials and consults as needed with the technical writer and/or product developers to obtain explanations of any ambiguous terms. The lead linguist also updates and validates the list systematically throughout the localization process.

Fonts

When it comes to fonts used in documentation, simpler is better. The conventional combination of a standard serif font (e.g., Times) for body copy and a standard sans serif font (e.g., Helvetica or Arial) for headings is a good example of font selections that work well for translation.

Try to use no more than three to four different fonts in a document.

Fonts such as Times, Helvetica, and Arial contain an extended character set that provides accented letters used in most Western European languages. However, Central European languages contain characters not included in the more commonly used fonts. If a document is targeted for Central European languages, in addition to other languages, it is important to choose fonts that have a matching CE versions such as Helvetica CE or Times CE. The localization vendor can help with the selection of an appropriate font. This is a good reason for determining ahead of time into which languages the document might be translated.

Asian Languages and separate operating systems

As mentioned above, many other languages require special fonts that are not available as extended character sets. Several even require separate operating systems. For example, Japanese, Korean, Traditional and Simplified Chinese are considered "double-byte languages." Each written character in these languages contains two bytes of data instead of 1 byte.

Character styles used in Western European or US English layouts do not always translate into Asian languages. In many cases, they are not used at all. For example, character styles such as bold and italic are not always applicable to Asian type styles and Asian characters do not distinguish between upper- and lowercase.

Layouts

White space

It is very important that the document's layout leaves enough room (i.e., white space) for the inevitable text expansion that occurs during the localization process.

As a general rule, allow for 20% to 30% expansion of your English text when it is translated. It is best to be conservative, using the 30% figure whenever possible.

Hyphenation

Hyphenation in the document also needs to be considered. The use of hyphenation affects the expansion of the translated text on a page. Non-hyphenated text generally takes up more space on a page due to the limited opportunity for convenient line breaks.

Page breaks

For technical support considerations, decide whether the translated documents should maintain the same page breaks and the same total of pages as the English document. It is generally easier if page breaks can change during the localization process. From the perspective of customer support, however, it is often preferable for the localized manuals to match the page breaks of the English so that support can always refer to "page 37 of the manual" for solving a problem. If page break matching is desired, it is even more important to allow for the "extra white space" described in **White space** (page 4).

When you hand off the document to a localization vendor for translation, always provide a hardcopy with the electronic files for double-checking.

Graphics

If text must be associated with a graphic, try to create the text as a separate component in the page-layout application used to create the document. A callout or a caption should ideally be a text block in the layout program., not an element of a graphics file. This makes it easier when translating the text in the layout program.

If you have to include text in the graphics files, leave it in text form, not formatted as a part of the graphic.

Translation of screen shot text is accomplished through localization of the software that was used to generate the English shots, followed by a regeneration of the screen shots using the localized software. When creating the screen shots in English, be sure to generate all of them at the same screen resolution and scale, saving the graphics file in the same file format used by the document layout application.

Help files

The same considerations of fonts, layouts, and graphics have to be considered when converting a printed document into an online help file.

When converting documentation to HTML format, the document headings convey the same structured sense of importance incorporated in the print document. A

document that does not use style tags efficiently (for example, uses a different style tag each time, to produce exactly the same formatting attributes), requires much more time to set than a document that uses only one style tag to represent this uniform style. Using consistent style definitions throughout your document allows HTML style tags to be generated in the localized files more easily. Therefore, be sure to use consistent templates and CSS files.

When translating help files, the localizer needs:

- Source content files (.RTF, .DOC, or .HTML), containing the bulk of the help file text,
- Project files (used to compile the help content)
- Bitmaps or other graphics that are used in the help files (often containing text that requires translation), and
- Other content files (such as .CNT file used to create the table of contents for the help file).

Adobe Acrobat PDF files

PDF files might be generated from the source documenation for use in distributing printable manuals. For a translator to be able to work with online documentation using Acrobat, they need the source files. The source file is translated and then converted into PDF afterwards. The PDF file then needs to be proofread for accuracy.

Analysis of available Help Tools

The information presented here was obtained from sales people at the respective companies, since the opportunity to test the creation of a help file in another language or the translation of help text was not available.

If the software product is on one of the New York Life Approved Software Tiers, that fact is indicated in the heading for that software product.

Doc-To-Help 6.0 – ComponentOne (Tier 1)

Localization capabilities of Doc-To-Help

Doc-To-Help 6.0 (D2H6) itself doesn't translate the Word files, but it can work with International versions of Word and support "double-byte character sets." A third-party translation company or person would have to translate the text in Word and the Word file reimported into the project. However, the button and tab labels in HTML Help, e.g., Contents and Index, are not changed when D2H6 generates the help file. This may be able to be changed in one of the HTML Help source files, but more research would be required to find this out for sure.

Features of Doc-To-Help

The following are features listed on ComponentOne's website for Doc-To-Help 6.0.

D2H6 works from a Word file as its source and creates help files and printed documentation from the Word file. Creating online help and printed documentation is straightforward and the resulting files come through fairly cleanly.

Single-source authoring

Create HTML Help, cross-platform HTML Help, WinHelp, JavaHelp™, and print documentation from a single set of Word documents – there's no need to purchase separate editions, as you do with RoboHelp. Because Word documents are used as the source, it is fairly easy to import data that has been written to date into the software.

Database architecture

D2H6's project architecture stores the relationships among styles, topics, index keywords, and table of contents entries in a per-project database, automatically

generating index entries, hyperlink tags, jumps and pop-ups, etc. You can also share information among multiple projects.

Flexible conditional text

Providing links to related topics in the online Help, but omitting them from your printed manual is no problem because, in addition to the benefits of single-source authoring, D2H6 provides the ability to designate portions of your source documents for individual platforms, Help targets, or attributes (author-defined criteria).

Margin notes

You can add margin notes by clicking the **Insert Margin Note** button on the D2H6 toolbar and entering the desired text. You can also make those margin notes appear in the online Help as pop-ups by clicking the **Link to Margin Note** button when the text is entered.

Glossary terms

An empty glossary is automatically prepared when you create a new project. Simply open the glossary document and enter text as desired. During compilation, project documents are searched for text that matches glossary entries and those matches are converted into hypertext links.

Enhanced indexing

As your source documents are compiled, topic hierarchy information and any existing XE field data are extracted and saved to the project database. You can then attach index keywords to topics and just push a button to rebuild the project and update its index.

Scripting

If you're a programmer, D2H6 allows you to use Microsoft Visual Basic® scripts to customize the behavior of individual styles at compile time. This feature has not been needed or used by the Agency help development team, but it might be useful to the programming team.

Modular help projects

Unlike multi-file projects, in which several Word document files in the same folder are compiled together into one Help file, modular Help projects consist of documents in different folders that are part of different Help projects, and are compiled into different Help files that appear as one file to users.

Contact and price information for Doc-To-Help

Website: http://www.componentone.com/cmd.boa?product=DTH6

(800) 858-2739 x233

Talked to Adam

$499 per copy – each user needs one copy

Telephone Support for ComponentOne Doc-To-Help 6.0 per year: $200

RoboHelp Office 2002r2 – eHelp Corporation (Tier 1)

Localization capabilities of RoboHelp

Currently, RoboHelp provides the capability of selecting one of 11 languages within the software. Asian Language and Japanese versions of RoboHelp are also available that support double-byte character sets.

RoboHelp can import Word files into the topics. It seems that when documentation is translated in Word files, the resulting files can be reimported into the standard version of RoboHelp.

Features of RoboHelp

The following are features listed on eHelp's website for RoboHelp Office 2002r2.

The project source is created in topic form and the target help file is generated from that. Both HTML Help and a web version of the help file can be created. The Web version is called **WebHelp** and creates a help system that looks and operates very similarly to HTML Help.

There are a lot of good features of RoboHelp for creating online help, but the software does not create printed documentation well. The resulting Word file may need tweaking to look the way you want it to.

Smart MS Word Import

Automatically separate a Word document into multiple Help topics based on styles and preserve the table of contents and index of the Word document with the Smart Word Import feature.

Office XP Word 2002 Support

You can develop Help projects with the newest version of Microsoft Word – Word 2002 in Office XP. Previous versions such as Word 2000, 97 and 95 continue to be supported.

Windows XP Support

WebHelp's Pure HTML and DHTML options overcome Microsoft's removal of the Java Virtual Machine on Windows XP. Windows ME, 2000, 98, 95 and NT 4.0 operating systems continue to be supported.

Output Multiple Formats

Create Help in all Help formats - WebHelp, Microsoft HTML Help, WinHelp, Oracle Help for Java and JavaHelp™ format from a single set of source code. Notice, however, that printed documentation is not listed. Generating printed documentation is not one of RoboHelp's strong points.

Supports Popular Editors

Develop all Help formats with any HTML editor, our integrated WYSIWYG HTML editor or Microsoft Word.

International Capabilities

Author in 11 different languages including; English, Danish, Dutch, French, German, Italian, Norwegian, Portuguese, Spanish, Swedish or Brazilian Portuguese. RoboHelp Asian Editions also available.

End users have full-text search capabilities in their native language.

Glossary Tab in WebHelp

The glossary in your WebHelp project makes it possible for end users to quickly look up and learn key terms.

Customizable Look and Feel

Customize the navigation panes, toolbar, icons, fonts, buttons and colors of your Help systems with Skins. Apply New York Life's look and feel to WebHelp projects. This feature has been used in the help files for FTIS Web, using New York Life's colors from the Intranet site and the New York Life logo.

Topic Templates

Topic templates let you save any topic and reuse it as the basis for new topics. Topic templates can include Headers and Footers (see **Topic Headers and Footers** below), cascading style sheets (CSS) and "seed" text. Templates save you time and effort during the development process while giving your Help systems a standardized look and feel across your company. This feature has been tried and found to be useful.

Topic Headers and Footers

Add Headers and Footers across topics, just like Microsoft Word, and add company logos, confidentiality information, copyrights, important Web site URLs and more, to each topic. Like Word, Headers and Footers are quickly updated by allowing a single change to be made to multiple topics at once.

XML Navigation

End users will experience significantly faster load times thanks to new XML-based navigation. XML is becoming the norm for structuring Help information. Microsoft uses it in MS Help 2.0 (the new version of their Help format). RoboHelp makes this technology available now, in its WebHelp Help format. The use of XML file formats allows the table of contents (TOC), index, search and glossary data to download on demand. Only the data that is required by the current inquiry downloads to the end users' browser. This removes the upfront delay of having the entire navigation files download when the user first accesses the Help system.

Spell Check of TOC, Index and Glossary

Find and correct embarrassing spelling errors in your navigation as well as in your Help topics.

Customizable Popups

You can manually size and override the automatic resizing of popup windows to control how they display for your end users. If the text goes beyond the height or width of the manually sized popup, scroll bars are provided.

Contact and price information for RoboHelp

Website: http://www.ehelp.com/products/robohelp/office/

(800) 358-9370

Talked to Adam

$999 per copy – each user needs one copy

Call for multi-user license prices.

eHelp also offers tech support programs – each is a 1-year subscription – that are detailed on their site. The plans start at $399 and go up to $999 per year.

Combining Doc-To-Help and RoboHelp

A combination of Doc-To-Help and RoboHelp is used to create the user guide help files for FTIS.

Doc-To-Help is used to maintain the source files. It is good for maintaining single-source documentation, in that it has a conditional text feature. When generating the target help files or printed documentation, you can define several targets. For example, two types of targets are created: HTML Help and Printed Documentation. For each of the two targets, an FTIS Desktop version and an FTIS Web version have been defined.

When HTML Help is generated for FTIS Web, the resulting help system is imported into RoboHelp. RoboHelp's WebHelp feature is used to generate help for FTIS Web that looks and operates similarly to the help files in FTIS Desktop.

If you are distributing the software and the help files on CD, you can generate HTML Help for now. When you get to the point where you are posting the software to a web server for access by the users, you can still generate HTML Help from Doc-To-Help and import the project files into RoboHelp and generate WebHelp for use on the web server.

Plain HTML can be generated from Doc-To-Help for use on a web server. A table of contents and an index are created, however it looks different from HTML Help and there is no search capability.

ComponentOne, the makers of Doc-To-Help, has bought the rights to parts of a help authoring tool from a company that went out of business, and one of the parts is called InterHelp. InterHelp is another version of WebHelp. It is a web-ready version of HTML Help. ComponentOne is working on incorporating InterHelp into Doc-To-Help, but they do not know at this time when that will happen and whether that will consitute a paid upgrade or a free upgrade.

AuthorIT 3.0

Localization capabilities of AuthorIT

AuthorIT provided the most succinct information about localizing documentation in their help authoring tool, AuthorIT. The localization capabilities of AuthorIT are more clearly defined than the information obtained about Doc-To-Help and RoboHelp.

AuthorIT currently supports all Western European single-byte character sets, but not yet double byte (at this stage). They are developing a Localization module within the next 6-12 months which will allow you to simultaneously store your documents in different languages in the same database. They are working with multiple localization vendors and the module will integrate with translation tools such as Trados.

They provided a recommendation for how to handle multiple language versions of the same document until the update is available – they referred to a specific Knowledge Base article called "Managing Multiple Languages of the Same Document."

Version 4 is due out later this year, and it will begin to implement support for double byte character sets (such as Chinese). Any data created in a library now, will be upgraded automatically when updating to Version 4. Full support will be implemented when the Localization module is released.

Features of AuthorIT

AuthorIT is a help authoring tool that has has the most flexibilty. It is a true single source documentation tool. Authors create pieces of reusable information. These pieces are managed and maintained in one database – the single source – then published to many audiences, documents, and formats – both print and online.

You can use and re-use information in as many places as you want. Re-use paragraphs, topics, sections, chapters, even whole documents. Re-use text, graphics, formatting, layout, linking, and indexing. Conditionally include and exclude information at any level. By re-using information, you write information once, make changes in one place, reduce errors and re-work, eliminate duplication, minimize reviews, reduce translation and localization costs, and maintain consistency.

An evaluation version of AuthorIT was tried out about a year ago. A smaller version of our FTIS help file was created and the software was found to be very useful and very flexible. However, switching to a new authoring tool couldn't be justified at the time.

Collaborative Authoring

AuthorIT is a collaborative authoring tool, where authors and other contributors work together as a team. AuthorIT provides a multi-user environment with dynamic check-in and check-out, security control, standards control, version control, document release control, and integrated project and task management.

Import Existing Documentation

AuthorIT imports documentation from Windows Help, HTML, XML, Microsoft Word, RTF and FrameMaker MIF documents.

No Tweaking or Rework required

When Doc-To-Help and RoboHelp is used to create the online help files for FTIS, some rework and tweaking is required to get the resulting files to look and behave the desired way. When AuthorIT was tried, all preferences were set in the project and the resulting help file and printed documentation was ready for use without extra tweaking or rework.

Every aspect of each format can be defined within the authoring environment, so your published documents are ready to go, with no additional tweaking or re-work. Print documents are print ready, with title pages, table of contents, headers and footers, pagination, section layouts, chapter contents, page number cross-references, margin notes, glossaries, and indexes. Online help and web formats include table of contents, indexes, full text search, customizable navigation, hypertext linking, popups, and more.

Full Text Search in HTML

You can include a full text search in the resulting HTML documents making information easier to find. AuthorIT provides features to allow combining any search engine with your HTML documents.

Contact and price information for AuthorIT

http://www.authorit.com

Phone number and reseller information is on their site. The main company is located in New Zealand, so calling them is tricky – I think it is 17 hours later there.

I have provided prices for both the Workgroup Edition and the Desktop Edition and for all levels of support plans. I recommend the Workgroup Edition for purposes of having multiple authors working on the project. See **Collaborative authoring** (page 12).

For more information about their support plans, see their website.

Workgroup Edition

With no support:	$1,497.00
With Standard support	$1,721.55
With Full support	$1,871.25

Desktop Edition

With no support:	$597.00
With Standard support	$686.55
With Full support	$746.25

Recommendations

Among the three tools mentioned in this document, AuthorIT is recommended. Doc-To-Help and RoboHelp are currently used because those are the tools available. If you have your choice of help authoring tools from the start, AuthorIT would be the best choice. There is a slightly higher learning curve to learn how to create documentation using the "reusable bits of information" process that AuthorIT uses, but it would be the best tool in the long run. It is the most flexible tool and will be the most useful. Their concept of creating reusable bits of information is the way a single-source help project should be handled to reduce errors, maintenance time, and duplication.

They seem to have a more thorough plan for incorporating localization capabilities than Doc-To-Help or RoboHelp. They don't specify specific languages as RoboHelp does. It sounds like once they have the Localization module completed, we will be able to produce documentation in whatever lanaguages we need and store the documentation in all the languages we need in one database. The time frame for their completion of the Localization module might fit in with your needs because, from the meeting we had on June 3, it sounds like you won't be translating the documentation for a while yet.

If you have to get a tool that is on the approved software list, Doc-To-Help is recommended. It is actually recommended that you combine Doc-To-Help with RoboHelp, but that may not be practical. The main reason to combine the two is so you can create WebHelp for future use on a web server, when you release the software on the web. Once ComponentOne releases an upgrade that includes Interhelp (their version of WebHelp) as mentioned earlier, RoboHelp may not be used as much. Currently, JavaHelp is available as a help format to generate, however, not enough is known about this format to recommend using it. The style of it looked rather strange when it was generated as a test and the process for setting up styles and stylesheets for JavaHelp is not known.

However, RoboHelp is also useful for tweaking the resulting help project from Doc-To-Help, which needs to sometimes be done to the FTIS help files. This can sometimes be accomplished by tweaking the code in the resulting project files using Notepad, but the extent to which tweaking can be done in this manner is not known (the programming capabilities are to limited on the FTIS help team to know). Doc-To-Help is closer to AuthorIT when it comes to true single-sourcing than RoboHelp. The HTML Help and Printed Documentation that Doc-To-Help creates is fairly clean.

Help Authoring Tool Training

If the desire is to have training on the help authoring tool in-house and AuthorIT is selected as the tool, training is available through some of the resellers that sell AuthorIT, and at least one of the ones looked at will come to our site to train you. They charge per day and allow up to 6 students in a class for the price. They will customize the class to train your people on how to use AuthorIT for your specific needs.

If RoboHelp or Doc-To-Help is chosen, training can be provided by someone in the Agency department.