

# Improving the FreeBSD Translation Tools

Warren Block <[wblock@FreeBSD.org](mailto:wblock@FreeBSD.org)>

## Abstract

Translation of FreeBSD's English documentation into other languages is extremely important to users and the FreeBSD project itself.

The old methods used to translate FreeBSD documentation are very labor-intensive. Existing translators have difficulty keeping translations up to date, new translators are discouraged by the depth of expertise required to translate, and even writers working on the English original documents must take extra steps to accommodate translations.

The `gettext` PO translation system automates much of the work and provides new abilities to make translation easier and faster, to share the work with other members of a translation team, and to reuse previous work. The PO translation tools have now been implemented for translating FreeBSD books and articles and this new method has begun to revitalize our translation efforts.

Here we describe the benefits of translation, the difficulties with the old method, the benefits of the new system, and some potential challenges and possibilities for the future.

## 1. Introduction: The Importance of Translation

Translations of FreeBSD documents into other languages are very valuable to FreeBSD. Translation lowers a huge barrier to entry and makes FreeBSD accessible to the vast group of people who do not speak English. That group benefits from access to FreeBSD, and FreeBSD benefits from the larger community.

The people who do the difficult technical work of translating FreeBSD documents into other languages accomplish more than just the translation. They act as ambassadors, spreading the word about the benefits of FreeBSD to previously-unserved groups, and enriching FreeBSD's user base. However, the valuable service that translators provide is hindered by the difficulty of the translation process. Bringing modern translation tools to FreeBSD benefits the translators, the FreeBSD community, and ultimately FreeBSD itself.

## 2. The FreeBSD Documentation

FreeBSD is famous for the quality of its documentation. This documentation consists mostly of three general categories: books and articles, manual pages, and the web site.

Books and articles are marked up with the [DocBook markup language](#)<sup>1</sup>.

The biggest and most important book is the [FreeBSD Handbook](#)<sup>2</sup>. This book and some of the other books and articles have been translated into more than twenty languages.

DocBook markup is XML-based and somewhat difficult for humans to use directly. Still, most of the authors and editors are comfortable enough with this code to edit it in a text editor. It works for us, but is intimidating for new users.

Manual pages are quick-reference documentation for most FreeBSD commands and configuration files. Manual page text is formatted with the [mdoc\(7\)](#)<sup>3</sup> markup language. At present, the FreeBSD manual pages have only been translated into [Japanese](#)<sup>4</sup>.

While simpler than DocBook, `mdoc` is still relatively complex and similarly intimidating to new users.

The FreeBSD web site<sup>5</sup> is produced by a combination of technologies. Many pages are XSLT transformations of XML files, while others are generated by CGI scripts or written by hand in raw HTML.

## 3. Whitespace

*Whitespace* is the name given to space separation characters in text. This includes spaces, tabs, carriage returns, and line feeds. All of these characters are mostly or entirely invisible. It is very difficult for translators to distinguish between whitespace-only changes that do not need to be retranslated, and actual content changes.

Accordingly, people working on the English original documents [must manually separate out any changes to whitespace](#)<sup>6</sup>. We cannot rewrap a paragraph or add a space before a sentence in the same commit with any content change, or the translators might have to retranslate sections that have already been translated.

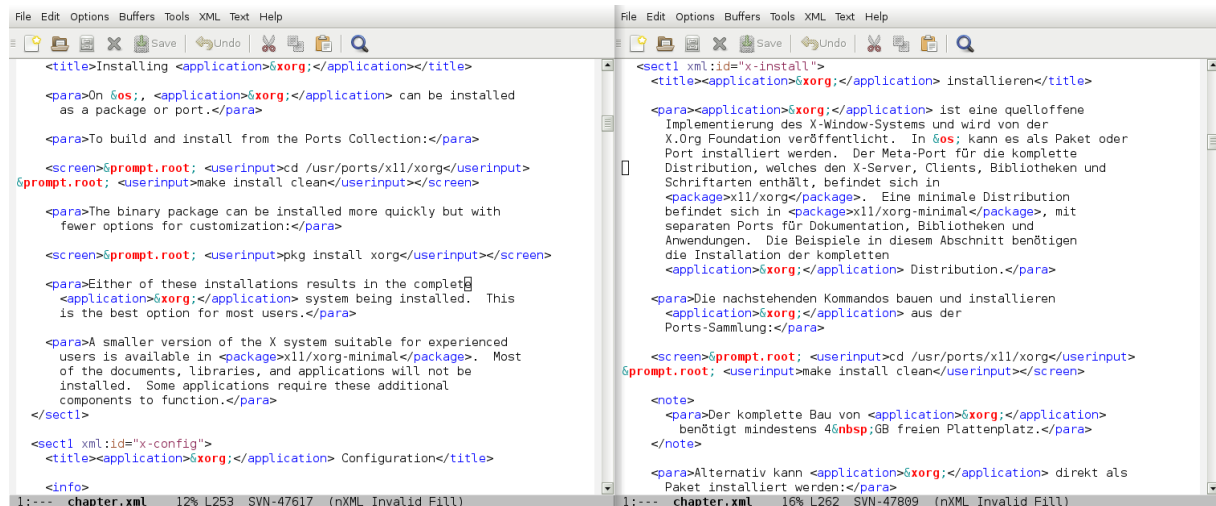


Figure 1: The old translation system. English document in editor window (left) and incomplete German translation (right). There is no one-to-one correspondence, and the translator must preserve all of the DocBook tags.

## 4. The Problem: The Old FreeBSD Translation Method

The old way of translating FreeBSD documents is not so much a cohesive system as a process that grew out of [manual document translations](#)<sup>7</sup>.

### The Old Translation Process

- The translator checks the comment at the top of a translated file, identifying the revision number of the English document with which it corresponds.
- The translator uses the revision control system to determine the number and order of changes that have been made to the English original since the revision number listed in the translated document.
- Each change to the English version is translated and manually entered into the translated version. The translator cannot prioritize changes or translate small pieces of a large change. They must be translated and added in the same order and size as the changes to the English original.
- As the translator works, they reformat the text according to the formatting rules of their translation team. These rules differ from the rules used in formatting the English version, and also differ between the various translation teams.
- When the translation of a particular revision is complete, the translator updates the

revision number at the top of the translated file and commits the new translation to the repository.

The requirement of translating changes in the same order in which they occurred in the English document is problematic. Translators might wish to translate a more important change before other changes. Because later changes are based on differences with earlier ones, this is difficult or impossible to do. Identifying the order and content of changes to the English original requires some in-depth knowledge of the revision control system.

Sections are often rewritten or heavily modified in the English version. Translators find themselves translating a large section only to have it deleted entirely in a later change.

Some words and sentences occur in many documents, but still must be retranslated for each new document. There is no way to automatically reuse these frequent translations, or to standardize on a single translated version for consistency.

Translation within these constraints is so much work that most translation teams struggle to keep up with the English originals. As volunteers, some find it takes too much time and work to do as a hobby, and abandon the effort. As a result, FreeBSD's famous documentation does not have current translations in several important world languages, or the existing translations are so old as to be nearly useless.

The learning curve of the old translation method also prevents new volunteers from contributing. We do have people offer to help with translations. But the sheer amount of education and effort required prevent many of them from contributing at all. It is just too much work.

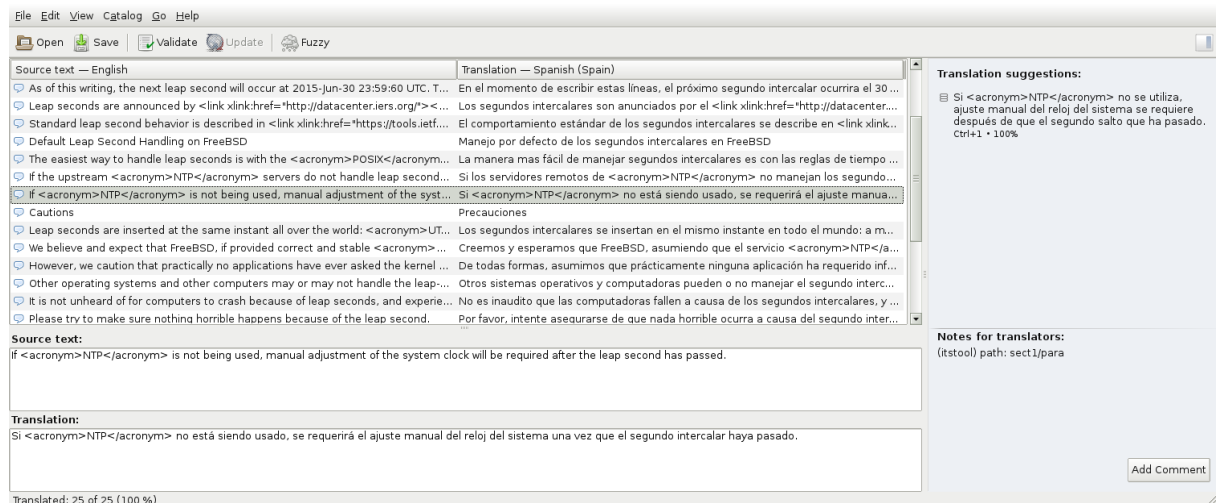


Figure 2: PO editor showing English and Spanish translations. Note the one-to-one correspondence and the lack of structural DocBook tags, which are handled transparently by the translation system.

## 5. A Solution: The PO Translation System

In the mid 1990s, a system for [internationalizing](#)<sup>8</sup> GNU programs developed into what is now called [gettext](#)<sup>9</sup>. `gettext` was originally meant only for use in programs. Each program contains translations of the text strings it can display. When the user has configured their [locale](#)<sup>10</sup>, the program displays the translated strings for that locale. In effect, it is a lookup substitution system using the English string as a key.

The `gettext` system also came to be used for documentation. A program extracts strings to be translated from a source file. These strings are stored in a [Portable Object \(PO\) file](#)<sup>11</sup>.

A *PO editor* program allows translators to enter a translation for each of the extracted strings. When complete, the original file is recreated with the translated strings replacing the English originals.

### The PO Translation Process

- The `make po` command is used to generate a PO file from the original English document. Translated strings already present in the PO file are preserved, so this command is also used to update the PO file when the original English document might have changed. Typically, this command is used once each translation session.
- A PO editor is used to enter translations for strings that have not yet been translated. The translator chooses which strings to translate. There are no restrictions on the

order of translation, nor is the translator forced to translate a specific group of strings that correspond to a single change in the English original.

- The updated PO file is saved and the `make tran` command is used to generate the translated document. This produces a version of the original book or article where the English strings have been replaced with the translated versions. This new file works just like the original, and can be built into HTML, PDF, or other output formats just like any other FreeBSD DocBook documentation.

There are several fundamental advantages to this system:

- `gettext` and PO translations are a widely-used standard.
- Whitespace changes are ignored by the system. Translators do not see them, and writers working on the English originals do not need to separate whitespace and content changes.
- Translators do not need to deal with DocBook structural tags like `<title>`, `<sect1>`, and `<para>`. These tags are handled automatically by the system and not even shown to the translator.
- Translators are not forced to monitor changes in the original file, translate them in order, or translate an entire commit in

one session. Translations are always based on the latest version of the English original document.

- The resulting translated document is generated by the system. Translators do not need to edit it to comply with any formatting rules.
- Translators only need to learn how to use a PO editor and some [basic information about DocBook and the revision control system](#)<sup>12</sup>.
- Multiple PO editors are available. There are online versions like [Pootle](#)<sup>13</sup> and [Weblate](#)<sup>14</sup>, so translators can work and collaborate without any software but a web browser.

But the PO system goes beyond this.

- Tools are available to extract source strings from more than one type of file. Of specific interest to FreeBSD are the tools to create PO files from man pages.
- Translated sections are added to Translation Memory (*TM*) and are suggested when text is re-used.
- Translation memory can be shared between translators. After a string has been translated once, other translators do not need to retranslate it.

## 6. Implementation

[itstool](#)<sup>15</sup> was chosen as the newest and most standards-compliant program to extract translatable strings from XML files and then to rebuild the translated documents. [po4a](#)<sup>16</sup> is another possibility. Although not as technically correct as *itstool*, the *po4a* package is compatible with many more types of translatable files.

The [msgmerge](#)<sup>17</sup> and [msgfmt](#)<sup>18</sup> utilities from the [devel/gettext-tools port](#)<sup>19</sup> are used for merging new translations with existing ones and creating the binary translation format.

Use of *itstool*, *msgmerge*, and *msgfmt* is integrated into the FreeBSD documentation Makefiles. Translators do not need to know the details of how these utilities or the FreeBSD documentation build system works. Most of the complexity is encapsulated by the simple `make po` and `make tran` commands.

## 7. Outcome

The initial integration of the PO system into the FreeBSD documentation build system was done in August, 2015. A new chapter describing the use of the PO system was added to the FreeBSD Documentation Project Primer in September, 2015.

Since then, eight new translations of articles into Spanish, Dutch, and Korean have been completed, and translations of entire books have begun.

## 8. Challenges and Potential

The original FreeBSD translation method evolved over a period of twenty years. At present, work is just beginning with the PO system. Both existing and new translators have used it successfully, and the feedback has been used to refine the process. This will continue as we discover new methods and procedures to make translation faster and more productive. FreeBSD might use one of the web-based PO editors for all the languages, [as PC-BSD has](#)<sup>20</sup>.

One or more translation teams might want to set up their own Pootle server, or use a crowd-sourced translation system like [Transifex](#)<sup>21</sup>.

There is also the question of whether the effort expended on existing translations can be reused with the PO system. [xml2po](#)<sup>22</sup> is an older program from the [GNOME documentation project](#)<sup>23</sup> which can generate a PO file when given an up-to-date translated file and the English original. [po4a-gettextize](#)<sup>16</sup> is a similar program from the *po4a* (PO for anything) project.

Lacking any way to associate the English and translated strings, both of these programs must compare the structure of the English files with those of the translated files. For this to work, the translations must be translations of the current English document. Unfortunately, very few of our translations are completely up to date, with a structure identical to the original English files. Where the original is composed of multiple files, like the chapters in a book, it might be possible to generate PO files from individual chapters that are up to date. For some documents, particularly shorter ones, it will likely be less work to just retranslate them with the PO system.

FreeBSD has other XML documents in addition to DocBook articles and books. Many of the project's web site HTML files are generated from XML source files. PO translations can be used with these files, although they are technically more challenging

because the layout is less regimented than that used in books and articles.

The PO system is not limited to XML documents. In fact, conversion programs are available to extract translatable strings from numerous other file types. Of particular interest is the ability to generate PO files for translation of manual pages. At present, the only translation of FreeBSD manual pages is the Japanese project. With PO translation tools, we could expand that support and make one of FreeBSD's greatest features available to non-English speakers around the world.

- Ryan Lortie
- Koop Mast
- Shaun McCance
- Chris Petrik
- Benedict Reuschling

## 9. Conclusion

The PO translation system provides a means for FreeBSD to accept the efforts of volunteer translators without requiring them to endure grueling workloads and manual processes. It eases the job of existing translators by automating many processes. Translation memory allows translators to share their work and not be forced to retranslate something that has already been translated once.

Translators gain an introduction to the FreeBSD documentation system, which in turn is an introduction to the ports and source systems that operate in the same way. This type of introduction frequently leads to new contributions in those other areas as well.

By lowering the barriers, FreeBSD documentation translations can be kept up to date and translated more easily. FreeBSD benefits from more exposure to the global community, and in turn, those people benefit from the option of using FreeBSD.

## 10. Acknowledgements

Bringing the PO translation system to FreeBSD was only possible with the support and help of a large group of people. Those who contributed most memorably are listed here, with apologies to the many others whose contributions were either too long ago or too subtle for this writer's memory.

- Thomas Abthorpe
- Glen Barber
- Mark Felder
- Hiroki Sato
- René Ladan

## Appendix A: References

- 1 DocBook Web Site: <http://docbook.org/>
- 2 The FreeBSD Handbook: [https://www.freebsd.org/doc/en\\_US.ISO8859-1/books/handbook/](https://www.freebsd.org/doc/en_US.ISO8859-1/books/handbook/)
- 3 mdoc Manual Page: <https://www.freebsd.org/cgi/man.cgi?query=mdoc>
- 4 Jpman Web Page: <http://www.jp.freebsd.org/man-jp/index.html>
- 5 The FreeBSD Web Site: <https://www.freebsd.org/>
- 6 The FreeBSD Documentation Project Primer: Whitespace Changes: [https://www.freebsd.org/doc/en\\_US.ISO8859-1/books/fdp-primer/writing-style-guide.html#writing-style-whitespace-changes](https://www.freebsd.org/doc/en_US.ISO8859-1/books/fdp-primer/writing-style-guide.html#writing-style-whitespace-changes)
- 7 The FreeBSD Documentation Project Primer: Translation (Old System): [https://www.freebsd.org/doc/en\\_US.ISO8859-1/books/fdp-primer/translations.html](https://www.freebsd.org/doc/en_US.ISO8859-1/books/fdp-primer/translations.html)
- 8 The FreeBSD Handbook: *Localization i18n/L10n Usage and Setup*: [https://www.freebsd.org/doc/en\\_US.ISO8859-1/books/handbook/l10n.html](https://www.freebsd.org/doc/en_US.ISO8859-1/books/handbook/l10n.html)
- 9 History of GNU gettext: [http://www.gnu.org/savannah-checkouts/gnu/gettext/manual/html\\_node/History.html](http://www.gnu.org/savannah-checkouts/gnu/gettext/manual/html_node/History.html)
- 10 GNU gettext utilities: *Locale Names*: [https://www.gnu.org/software/gettext/manual/html\\_node/Locale-Names.html](https://www.gnu.org/software/gettext/manual/html_node/Locale-Names.html)
- 11 Pology: *The PO Format*: [http://pology.nedohodnik.net/doc/user/en\\_US/ch-poformat.html](http://pology.nedohodnik.net/doc/user/en_US/ch-poformat.html)
- 12 The FreeBSD Documentation Project Primer: *PO Translations*: [https://www.freebsd.org/doc/en\\_US.ISO8859-1/books/fdp-primer/po-translations.html](https://www.freebsd.org/doc/en_US.ISO8859-1/books/fdp-primer/po-translations.html)
- 13 Pootle Web Page: <http://pootle.translatehouse.org/>
- 14 Weblate Web Page: <https://weblate.org/en/>
- 15 itstool Web Site: <http://itstool.org/>
- 16 po4a Web Site: <https://po4a.alioth.debian.org/>
- 17 msgmerge Manual Page: <https://www.freebsd.org/cgi/man.cgi?query=msgmerge&manpath=FreeBSD+10.2-RELEASE+and+Ports>
- 18 msgfmt Manual Page: <https://www.freebsd.org/cgi/man.cgi?query=msgfmt&manpath=FreeBSD+10.2-RELEASE+and+Ports>
- 19 devel/gettext-tools Port: <http://www.freshports.org/devel/gettext-tools/>
- 20 PC-BSD Pootle Web Page: <http://translate.pcbsd.org/>
- 21 Transifex Web Page: <https://www.transifex.com/>
- 22 xml2po git Repository: <https://git.gnome.org/browse/gnome-doc-utils/tree/xml2po>
- 23 GNOME Help Web Site: <https://help.gnome.org/>