Improving translation workflow for FreeBSD Documentation with Zanata

FreeBSD DevSummit @ BSDCan 2019

Who are we?



Edson Brandi (ebrandi)

- CIO at StoneCo Ltd (STNE)
- Founder of Brazilian FreeBSD User Group (1995)
- FreeBSD User since 1995
- FreeBSD Committer since 2012



Danilo G. Baio (dbaio)

- Sysadmin/IT Manager at CRMall/BS2
- FreeBSD Committer since 2017



Guangyuan Yang (ygy)

- Undergraduate Student at the University of Waterloo
- FreeBSD Foundation Co-op (Spring 2017)
- FreeBSD Committer since 2017

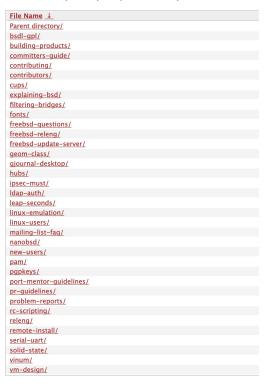
Agenda

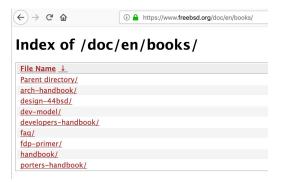
- Documentation Project Overview
- Why we need better tooling
- Introduction to Zanata and transpopy
- Proposed new workflow for translators
- Discussions

Overview: FreeBSD Documentation Set



Index of /doc/en/articles/





35 articles / 08 books*

1.383,832 words

5.535 hours of work or 230 full days (estimation of zanata)

Handbook: 255.500 words / 1022 hours

Porters Handbook: 77.060 words / 308 hours

FAQ: 27.300 words / 109 hours

FreeBSD Releng: 4.329 words / 17 hours Leap Seconds: 538 words / 2.16 hours

•••

This is the challenge for new Languages that aim to translate the FreeBSD Documentation.

2	×	Case sensitive Exact m	atch			Stats by:	Messages Downloa	d files (zip)
Path	Document	Statistic	Complete	Incomplete	Remaining hours	Last Upload	Last Translated	Actions
oooks/	handbook	0%	0	255500	1022	03/05/19 21:00 by admin		.po
ooks/	porters-handbook	0%	0	77060	308.24	03/05/19 21:01 by admin		.po
oooks/	arch-handbook	0%	0	70178	280.72	03/05/19 20:58 by admin		.po
ooks/	developers-handbook	0%	0	58812	235.25	03/05/19 20:59 by admin		.po
ooks/	fdp-primer	0%	0	28610	114.44	03/05/19 20:59 by admin		.po
oooks/	faq	0%	0	27326	109.31	03/05/19 20:59 by admin		.po
articles/	committers-guide	0%	0	20164	80.66	03/05/19 20:57 by admin		.po
articles/	linux-emulation	0%	0	12499	50	03/05/19 20:58 by admin		.po
articles/	contributors	0%	13	11756	47	03/05/19 20:57 by admin	03/05/19 23:07 by admin	.po
ooks/	design-44bsd	0%	0	11097	44.39	03/05/19 20:58 by admin		.po
rticles/	serial-uart	0%	0	9487	37.95	03/05/19 20:58 by admin		.po
ooks/	dev-model	0%	0	9261	37.05	03/05/19 20:58 by admin		.po
rticles/	vm-design	0%	0	6712	26.85	03/05/19 20:58 by admin		.po
rticles/	rc-scripting	0%	0	6317	25.27	03/05/19 20:58 by admin		.po
articles/	pam	0%	0	6211	24.85	03/05/19 20:58 by admin		.po
articles/	vinum	0%	0	5969	23.88	03/05/19 20:58 by admin		.po
rticles/	contributing	0%	0	5566	22.27	03/05/19 20:57 by admin		.po
rticles/	problem-reports	0%	0	4625	18.5	03/05/19 20:58 by admin		.po
rticles/	freebsd-releng	0%	0	4329	17.32	03/05/19 20:57 by admin		.po
articles/	ldap-auth	0%	0	4157	16.63	03/05/19 20:58 by admin		.po
rticles/	freebsd-update-server	0%	0	4007	16.03	03/05/19 20:57 by admin		.po
rticles/	releng	0%	0	3829	15.32	03/05/19 20:58 by admin		.po
rticles/	bsdl-gpl	0%	0	3756	15.03	03/05/19 20:56 by admin		.po
rticles/	building-products	0%	0	3432	13.73	03/05/19 20:57 by admin		.po
rticles/	new-users	0%	0	3351	13.41	03/05/19 20:58 by admin		.po

The truth

The current translation workflow is very boring and discouraging.

How to make it more enjoyable?

How do we better motivate volunteers?





The problem with traditional translation workflow

Just to point out some of these "problems":

- Need to learn the Documentation structure
- Deal with Docbook, Subversion, Builds
- Coordinate translations in a external repository
- Each language team has their own workflow
- Tracking updates is hard
- etc.

All of these difficulties results in a lot of outdated translations!

The Solution: The PO Translation System

- PO translation support was added in Aug 24, 2015.
- You still need to know about the Documentation structure, Subversion, etc.
- The adoption of this new system is still slow
- Teams are still separated
- No source files (POT) for English documents.
- People that manages to translate a PO file, still needs to push it upstream.
 (Remember the size of the documents and in the ONE person working)
- Difficult to evolve without a coordinator

Number of PO files in the Doc tree

Although it was introduced more than 3 years ago, the system is still used by few languages...

```
$ find . -name "*.po" | sed s/\.// | awk -F \/ '{ print $2 }' | sort -n | uniq -c
3 de_DE.IS08859-1
8 es_ES.IS08859-1
1 ja_JP.eucJP
3 ko_KR.UTF-8
5 nl_NL.IS08859-1
38 pt_BR.IS08859-1
1 zh_CN.UTF-8
6 zh_TW.UTF-8
```

GitHub Workflow is not good enough

```
Add initial translation to Handbook: Preface #12
№ Merged
                                                                                                 Diff settings ▼
                                                                                                                 Review changes
           Changes from all commits ▼ File filter... ▼ Jump to... ▼
 ...
 394
        394
               #. (itstool) path: listitem/para
               #: book.translate.xml:1046
                msgid "<xref linkend=\"ppp-and-slip\"/> has been substantially updated."
             - msgstr ""
        397 + msqstr "<xref linkend=\"ppp-and-slip\"/> 包含了大量更新。"
 398
               #. (itstool) path: listitem/para
 400
       400
                #: book.translate.xml:1051
 401
                msgid "Many new sections have been added to <xref linkend=\"advanced-networking\"/>."
             - msgstr ""
        402 + msgstr "许多新的内容被添加到 <xref linkend=\"advanced-networking\"/> 中。"
 403
        403
 404
               #. (itstool) path: listitem/para
        404
 405
       405
                #: book.translate.xml:1055
 406
                msqid "<xref linkend=\"mail\"/> has been expanded to include more information about configuring <application>sendmail
             - msastr ""
        407 + msgstr "<xref linkend=\"mail\"/> 增加了关于配置 <application>sendmail</application> 的信息。"
 408
        408
```

GitHub Workflow is not good enough

翻译流程 Workflow

- 和 ygy 联系分配任务, 并更新状态。
- 从上游拉取最新数据:
 git pull --rebase upstream master
 若本地没有未保存的更改,而 rebase 出现了错误,可以尝试强制更新 repo:
 git fetch upstream master
 git reset --hard upstream/master
- 使用一种 PO 编辑器 (例如 Poedit Poedit 下载) 对认领部分进行翻译: poedit zh_CN.po

GitHub Workflow is not good enough

books/handbook Status

Status

Title	Assignee	Status	Pull Request
Preface	уду	Updated	#12
I. Getting Started 1. Introduction		Needs Update	
2. Installing FreeBSD		Needs Update	
3. FreeBSD Basics		Needs Update	
4. Installing Applications: Packages and Ports		■ Needs Update	
5. The X Window System		Needs Update	

Improving our translation workflow using Zanata

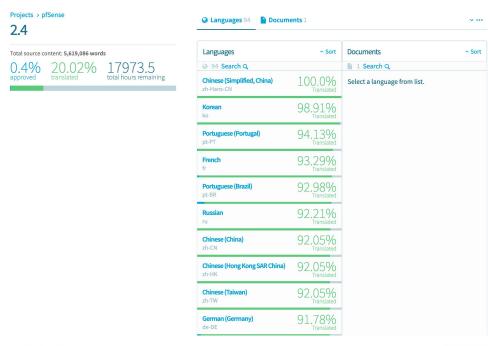
The idea came after seeing the success of pfSense translation.

In almost **2 years of work**, **20 languages** were translated (12 committed), a collective effort of **more than 400 volunteers**.

This is not a drastic change, *it's just a tool to gather all people that are willing to translate the FreeBSD Documentation* in the same place.

The **only prerequisite** for using it **is the PO system**, and it is already in our tree!

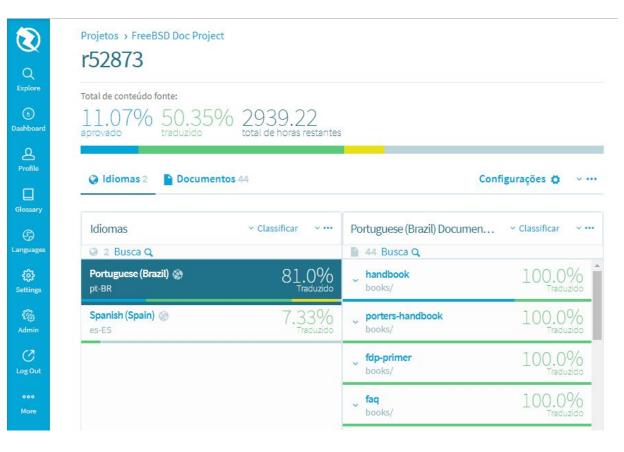
pfSense Zanata Project





Zanata use case for FreeBSD

Our Current Results



pt_BR:

9 months of work, 35 articles and 5 books translated, including The FreeBSD Handbook, by 40+ volunteers.

es_ES:

5 months of work, 18 articles by 2 volunteers.

zh_CN:

Getting started...

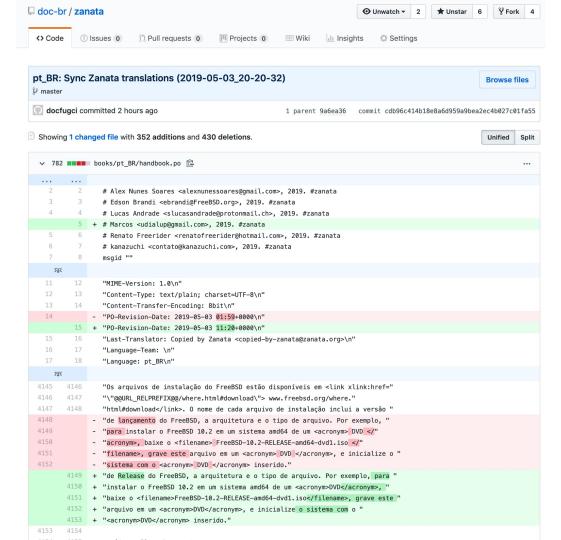
Benefits for FreeBSD provided by a web translation tool

- More contributors (no solo translators)
- FreeBSD Documents translated into more languages
- Lead time smaller to translate documents
- Easy to review and update translations (improve of quality)
- Have less outdated translations
- ...

Zanata Github Sync

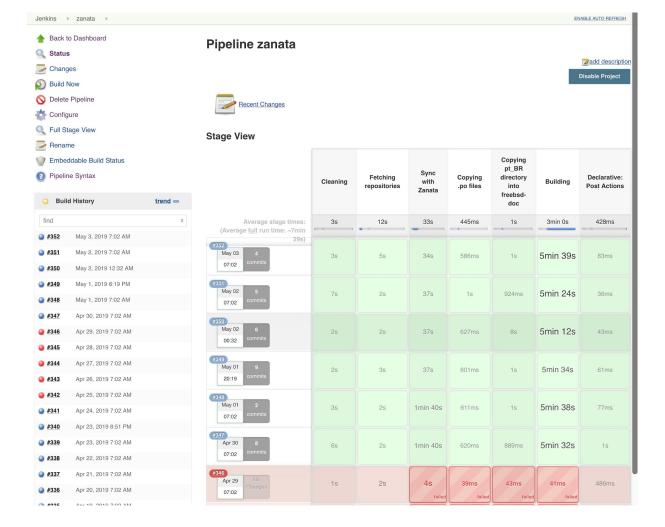
We do keep the documents being worked on the Zanata synchronized with a git repository in Github.

This helps us to keep a backup copy and to have visibility of the progress and the quality of the translations received.



Jenkins Automation

Using Jenkins to generate automatic builds makes it easy for volunteers to track the results and helps us monitor any code breaks.





Index of /jenkins

Name	Last modifi	ed	Size	
Parent Directory			_	
gh-290/	2019-05-03	22.24	_	
gh-289/	2019-05-03		_	
gh-288/	2019-05-03		_	
gh-287/	2019-05-03		_	
zn-352/	2019-05-03		_	
zn-es ES-96/	2019-05-03		_	
gh-286/	2019-05-02		_	
gh-285/	2019-05-02			
gh-284/	2019-05-02		_	
gh-283/	2019-05-02		_	
zn-351/	2019-05-02		_	
gh-282/	2019-05-02		_	
zn-es ES-95/	2019-05-02			
gh-281/	2019-05-02		_	
gh-280/	2019-05-02		-	
gh-279/	2019-05-02		_	
gh-2797 gh-278/	2019-05-02		=	
zn-350/	2019-05-02		_	
gh-277/	2019-05-02		-	
gh-277/	2019-05-02		_	
gh-275/	2019-05-02		0	
gh-274/	2019-05-01		_	
	2019-05-01		-	
gh-273/ gh-272/	2019-05-01		_	
	2019-05-01		-	
gh-271/ zn-349/	2019-05-01		_	
	2019-05-01		_	
gh-270/ gh-269/	2019-05-01		-	
	2019-05-01		-	
gh-268/	2019-05-01		-	
gh-267/	2019-05-01		_	
gh-266/ zn-348/	2019-05-01		-	
qh-265/			-	
zn-es ES-94/	2019-05-01		-	
	2019-05-01		-	
gh-264/	2019-04-30		-	
zn-347/			-	
gh-263/	2019-04-30		-	
zn-es_ES-93/	2019-04-30		-	
gh-262/	2019-04-29		-	
gh-261/	2019-04-29		-	
gh-260/	2019-04-29		_	
gh-259/	2019-04-29		-	
gh-258/	2019-04-29		-	
gh-257/	2019-04-29		-	
gh-256/	2019-04-29		-	
robots.txt	2018-11-01	20:32	26	

Jenkins Automatic builds



FreeBSD Handbook

Projeto de Documentação do FreeBSD

Revisão: d994e9dd5a

Copyright © 1995-2019 The FreeBSD Documentation Project

Copyright Nota Legal

2013-11-13 07:52:45 +0000 por hrs.

Resumo

Bem vindo ao FreeBSD! Este manual cobre a instalação e o uso diário do *FreeBSD 12.0-RELEASE* e do *FreeBSD 11.2-RELEASE*. Este livro é o resultado do trabalho contínuo de muitas pessoas. Algumas seções podem estar desatualizadas. Os interessados em ajudar a atualizar e expandir este documento devem enviar e-mails para a lista de discussão do projeto de documentação do FreeBSD.

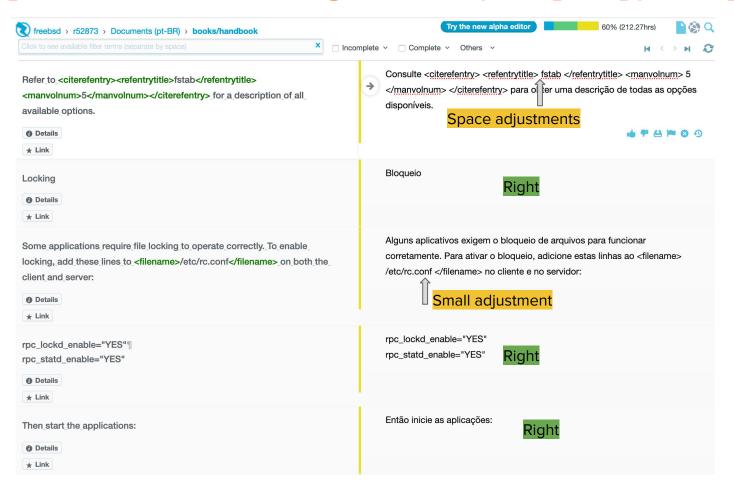
A última versão deste livro está disponível no <u>site do FreeBSD</u>. Versões anteriores podem ser obtidas em https://docs.FreeBSD.org/doc/. O livro pode ser baixado em uma variedade de formatos e opções de compressão do <u>servidor FTP do FreeBSD</u> ou de um dos inúmeros <u>sites espelho</u>. Cópias impressas podem ser adquiridas da <u>FreeBSD Mall</u>. As pesquisas podem ser realizadas no manual e em outros documentos na <u>página de busca</u>.

Speeding up the translation process with the Google Cloud Translation API

- You can generate drafts with textproc/py-transpopy (Created by kanazuchi, a Brazilian contributor)
- You need a Google Cloud Platform Account
- You need credits \$\$\$, but Any new GCP user gets 300 USD to try Google Cloud
- To generate a draft translation of all articles and books we used around 180 USD per language

Translation API version	Feature	1-500,000 characters	500,000 −1 billion characters
Translation API v2, v3	Language detection	\$20 per 1,000,000 characters*	\$20 per 1,000,000 characters*
Translation API v2	Text translation	\$20 per million characters	\$20 per million characters
Translation API v3	Text translation	Free*	\$20 per million characters
Translation API v3	Text translation (AutoML Models)	Free*	\$80 per million characters

Example of a draft translation generated by textproc/py-transpopy



What helped us to get more volunteers?

The award for the most committed volunteers was by far the initiative that brought the best results, the award takes place in the form of the distribution of gifts to the volunteers who contribute most to the translation effort.







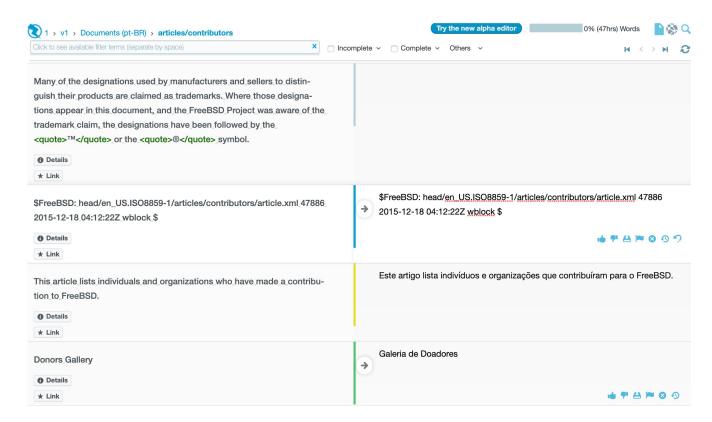




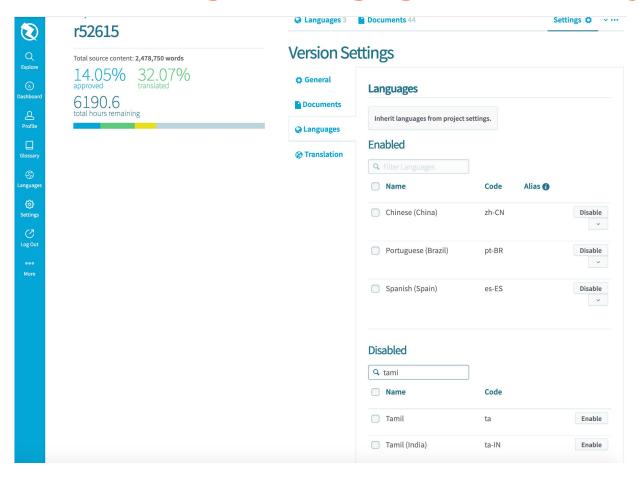
Zanata Overview

Zanata Translate Screen

The original text is displayed on the left and the text translated on the right.

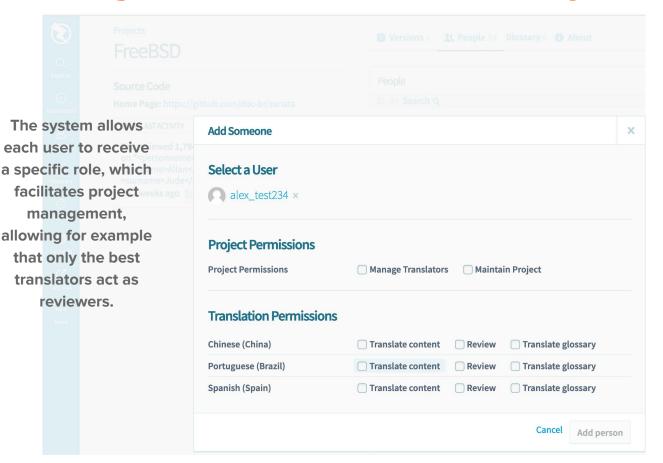


Enabling a new language in a Zanata Project



The project
administrator can
add a new language
for translation with a
few clicks.

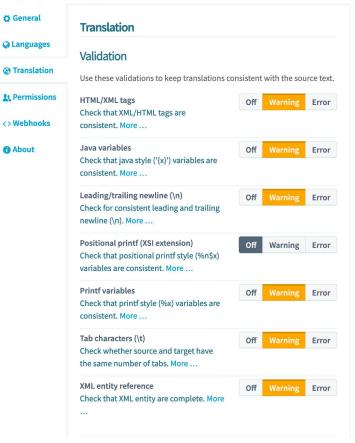
Adding a new contributor to a Zanata Project



Zanata Rules for Translation Validations

Project Settings

The rules system is very flexible and allows project administrators to have full control over them.



* Link If the translator <varname>DISTVERSION <varname> <varname>DISTVERSION</varname> makes a mistake with \rightarrow Details the tags, the zanata * Link will display an alert... Missing tag: </varname> Unexpected tag: <varname> <varname>PORTVERSION</varname> <varname>PORTVERSION</varname> Details * Link

Examples of <varname>DISTVERSION</varname> and the Derived <varname>PORTVERSION</varname> Details And when the error * Link was corrected the alert disappears <varname>DISTVERSION Details * Link

Link

Examples_of_<varname>DISTVERSION</varname>_and_the_Derived_

<varname>PORTVERSION</varname>

Details

Exemplos de <varname>DISTVERSION</varname> e de Derivações <varname>PORTVERSION</varname>. <varname>DISTVERSION</varname>

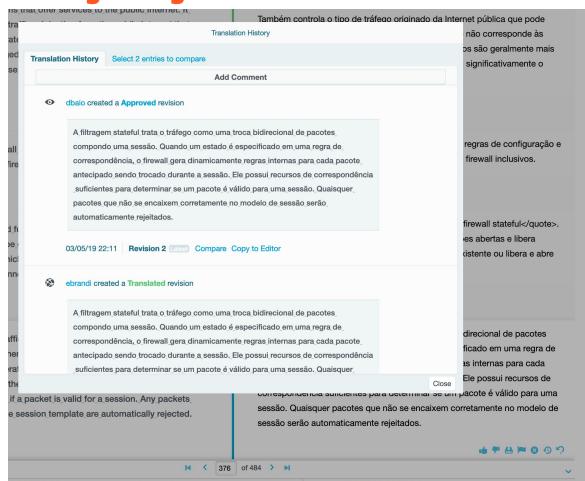
Exemplos de <varname>DISTVERSION</varname> e de Derivações

<varname>PORTVERSION</varname>.

 \rightarrow P 4 P 8 9 <varname>PORTVERSION</varname> <varname>PORTVERSION</varname> Details

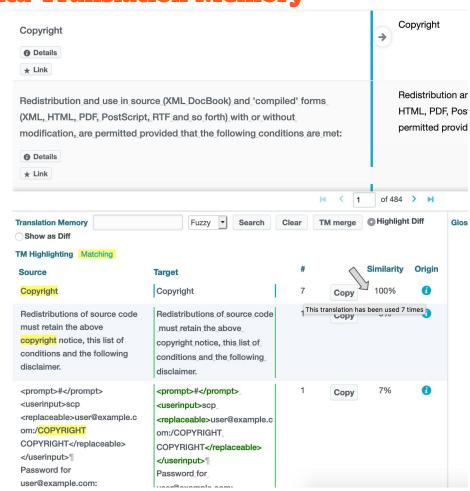
Tracking changes in translations

The zanata has its
own versioning
system, which allows
you to have a greater
control and
monitoring of the
translation
contributions made
by the volunteers.

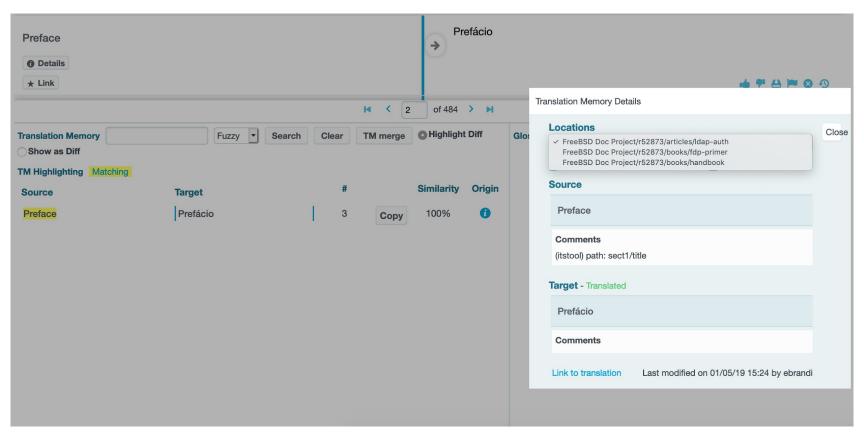


Zanata Translation Memory

The zanata has a memory system for translations, which streamlines and ensures consistency for the translation process.

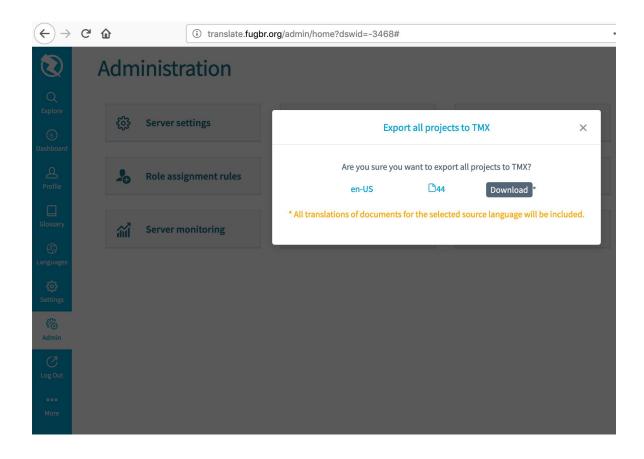


Zanata Translation Memory



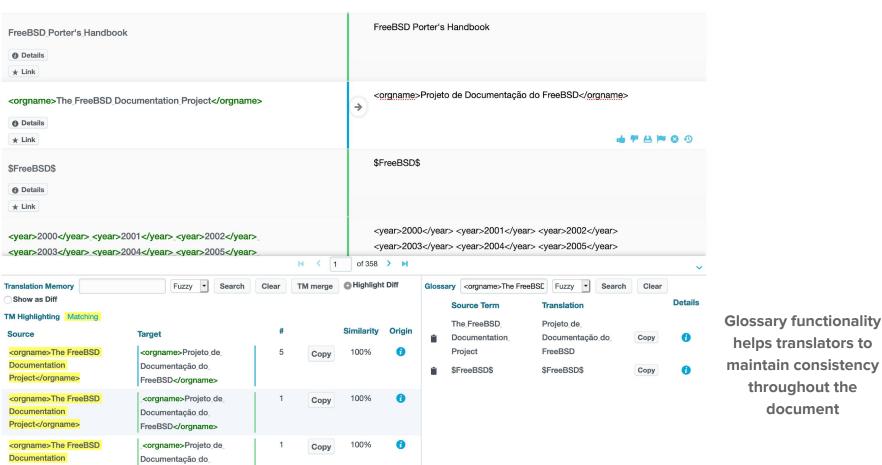
In this example we can see that the word "Preface" appeared in 3 other documents.

Export Translation Memory (Translation Memory Exchange)



Zanata allows the administrator to export the translation memories.

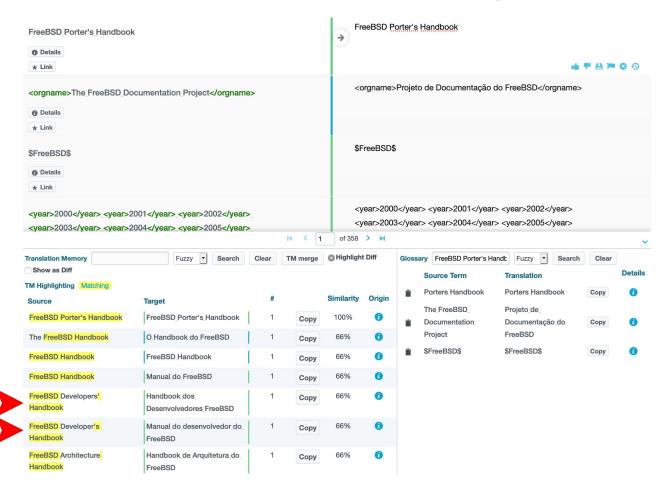
Zanata Glossary



Project</orgname>

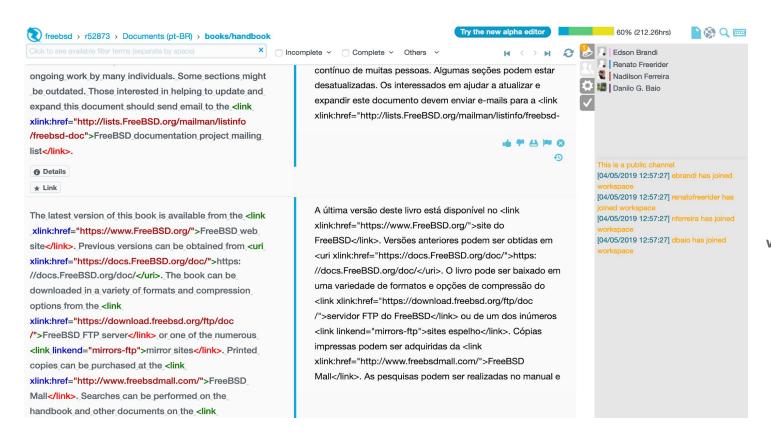
FreeBSD</orgname>

Zanata Glossary



The glossary feature also helps us identify situations where a term has received different translations.

Zanata Chat room



We can see who is working in the document and chat with them if we want.

Zanata Client

Push source content to the Zanata server.

Pull translated content from the Zanata server.

Port devel/zanata-cli

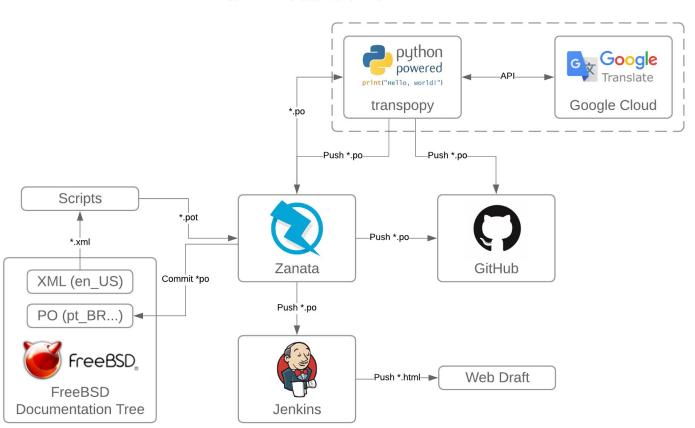
this tool helps us in many places

```
a]$ zanata-cli push --push-type source
[INFO] Loading project config from zanata.xml
[INFO] Loading user config from /home/dbaio/.config/zanata.ini
[INFO] client API version: 4.6.2, server API version: 4.6.2
[INFO] Server: https://translate.zanata.org/
[INFO] Project: FreeBSD
[INFO] Version: r52873
[INFO] Username: dbaio
[INFO] Project type: gettext
[INFO] Source language: en-US
[INFO] Copy previous translations: false
[WARN] As of Zanata Client 3.8.0, copyTrans is disabled by default.
[INFO] Merge type: AUTO
[INFO] Enable modules: false
[INFO] Include patterns:
[INFO] Exclude patterns:
[INFO] Case sensitive: true
[INFO] Default excludes: true
[INFO] Exclude locale filenames: true
[INFO] Pushing source documents only
[INFO] Current directory: /usr/home/dbaio/FreeBSD/github/doc-br/zanata
[INFO] Source directory (originals):
[INFO] Found source documents:
[INFO]
                  articles/bsdl-apl
                  articles/building-products
                 articles/committers-guide
                 articles/contributing
                 articles/contributors
                 articles/cups
INFO
                 articles/explaining-bsd
TINFOT
                  articles/filtering-bridges
                  articles/fonts
                 articles/freebsd-auestions
                 articles/freebsd-relena
                 articles/freebsd-update-server
                 articles/geom-class
INFO
                 articles/gjournal-desktop
INFO
                  articles/hubs
                  articles/ipsec-must
                 articles/ldap-auth
                 articles/leap-seconds
                 articles/linux-emulation
                 articles/linux-users
[INFO]
                 articles/mailina-list-faa
INFO
                 articles/nanobsd
                  articles/new-users
                 articles/pam
[INFO]
                 articles/pgpkeys
                 articles/port-mentor-quidelines
                 articles/pr-guidelines
[INFO]
                 articles/problem-reports
[INFO]
                 articles/remote-install
[INFO]
                 articles/serial-uart
                 articles/solid-state
                 articles/vinum
[INFO]
                 articles/vm-design
TINFOT
                 books/arch-handbook
                 books/design-44bsd
                 books/dev-model
[INFO]
                 books/developers-handbook
                 books/faq
                 books/fdp-primer
INFO
                 books/handbook
INFOT
                 books/pmake
                 books/porters-handbook
 ?] This will overwrite existing source documents on the server, and delete obsolete documents.
 re you sure (v/n)?
```

Proposed Workflow

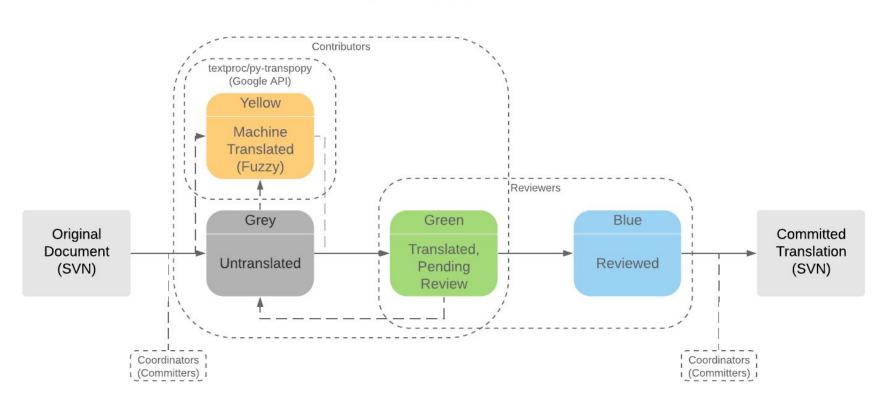
Translation Service Blueprint

Guangyuan Yang (ygy@) | May 16, 2019



Translation "States"

Guangyuan Yang (ygy@) | May 16, 2019



Proposed Workflow: Resources

- Zanata Web Application - https://translate.freebsd.org

- Zanata Repository - https://github.com/freebsd/zanata or svnweb.freebsd.org/doc/zanata?

- Bugzilla - New option for PR

Product: Documentation

Component: Translation (New)

Assignee: freebsd-translators@freebsd.org

- Jenkins - ci.freebsd.org

Jobs to build each language daily, the output need to be hosted somewhere like https://drafts.translate.freebsd.org

Proposed Workflow: Languages

Languages already using Zanata:

- pt_BR (Coordinator: ebrandi, dbaio)
- es_ES (Coordinator: gabor)
- zh_CN (Coordinator: ygy)

We can already add other languages, or just when we found coordinators/contributors

There are some languages like ja_JP and de_DE that don't use POT/PO files. We think it's better to not add these languages to Zanata or just add if the current translators asked for it.

Proposed Workflow: Roles

Reviewers

Reviewers will be only FreeBSD committers?

Contributors who show good work could become reviewers?

Coordinators

Coordinators can add new users to the language.

Coordinators can set users as translators or reviewers.

Coordinators will be only FreeBSD committers?

Contributors who show good work could become coordinators?

Contributors

For languages with contributors and no coordinators, we can add zanata-admin as coordinator

Proposed Workflow: New Users

For new users

Let any people register in the platform.

They will receive a "user" role and then they will need to ask to join in some language team as a "contributor".

These requests will be sent by email to the language coordinator. (all handled by Zanata)

Proposed Workflow: Commit of Translations

Translation with what status can be committed?

100% translated (green)

or

100% reviewed (blue)

Do reviewers explicitly need to mark all the text as reviewed (blue) in the Zanata interface?

We suggest reviewing all documents before committing it to the doc repository.

When a document is translated by a contributor and there is no coordinator in the language, the contributor can create a PR (Bugzilla) asking to add/modify the translated document to the doc tree. (Bugzilla → Documentation → Translation)

Credits to all the people in the PO file

Add them to the credits field in the PO file

Add them to contributors (head/en_US.ISO8859-1/articles/contributors)

Proposed Workflow: Update Zanata Source Files

Update POT files

- Monthly or Quarterly?
- Send an alert to the freebsd-translators@ mailing list?
- Register the update procedure in Wiki so any doc committer can do this?

Next Steps

PoC Zanata → https://translate-dev.freebsd.org
We can move pt_BR to there for a few months
Configure all resources (Jenkins/Output translations/Repository/Maillist/Wiki)

Zanata Public → https://translate.freebsd.org
Announce

Future:

Translate htdocs with Gettext (PO files) and Zanata ???

Translate man pages with Gettext (PO files) and Zanata ???