

Emacs manuals translation and OmegaT

Translation in the Free Software World

- Localization and translation are all over free software
- The GNU project warmly welcomes translation
- Tools and processes vary, but all the communities have found a way to provide localized versions of the software package they manage.

All the communities?

errr, well, not quite...

- GNU's own flagship, Emacs, is not there yet...
- There are plenty of *good* technical and human reasons why that is not the case (check the emacs-devel list for background information)
- But in the likely case that Emacs won't be fully localized before global temperatures rise by more than 2 degrees, we can still translate the manuals...

Emacs manuals

- as of today (nov. 7), wc output:
 - 182 files
 - 2,266,158 words
 - 15,830,468 characters
- 2000 words a day, 365 days a year = about 3 years, 1 month, 8 days...

Is everything good to translate?

- Obviously, not...

```
@c Merge all indexes into a single index for now.  
@c We can always separate them later into two or more as needed.
```

```
@syncodeindex vr cp  
@syncodeindex fn cp  
@syncodeindex ky cp  
@syncodeindex pg cp  
@syncodeindex tp cp
```

```
@c @footnotestyle separate  
@c @paragraphindent 2  
@c @@smallbook  
@c %**end of header
```

Only this part...

```
@copying  
This manual documents the Semantic library and utilities.
```



Formats!!! How to deal with "meta" things...

- Translators should not have to deal with that many "meta" things.
- Nobody would think of translating *inside* TeX, or DocBook, or, would you?...
- There is a way to deal with "meta" document things, just like there is a way to deal with "meta" code things.
- The `gettext` utility has brought us the Portable Object format for code
- `po4a` does the same for documentation

.texi / .org ↔ po4a ↔ P0

- Et voilà!

```
#. type: top
#: /Users/suzume/Documents/Repositories/Projet OmegaT de Documentation Emacs -
#: Sources/doc/misc/semantic.texi:5 /Users/suzume/Documents/Repositories/Projet
#: OmegaT de Documentation Emacs - Sources/doc/misc/semantic.texi:81
#, no-wrap
msgid "@value{TITLE}"
msgstr ""

#. type: include
#: /Users/suzume/Documents/Repositories/Projet OmegaT de Documentation Emacs -
#: Sources/doc/misc/semantic.texi:6
#, no-wrap
msgid "docstyle.texi"
msgstr ""

#. type: copying
#: /Users/suzume/Documents/Repositories/Projet OmegaT de Documentation Emacs -
#: Sources/doc/misc/semantic.texi:27
msgid "This manual documents the Semantic library and utilities."
msgstr ""
```

Not perfect,
but much better...

How many words now?

- OmegaT statistics on the same file set:
 - 182 files
 - 1,920,945 words (-15.2%)
 - 12,094,820 characters (-23.6%: *that much less to type!*)
- still quite a mouthful: about 2 years, 7 months, 20 days...

"OmegaT"? Did you say "OmegaT"?

- OmegaT is a GPL3+ Java8+ "Computer Aided Translation" tool
- CATs are not good only for cute pictures
- They are also good for professional translators
- CATs are to translators what IDEs are to programmers
- They leverage the power of the computer to automate reference searches

19 years old this month...

Downloads

1,375,629

2002-11-28 to 2021-11-07

Countries

Top*: **ES**, at 10%

Operating Systems

Top*: **Windows**, at 73%

Download Statistics

[All Files](#) ([Change File](#))

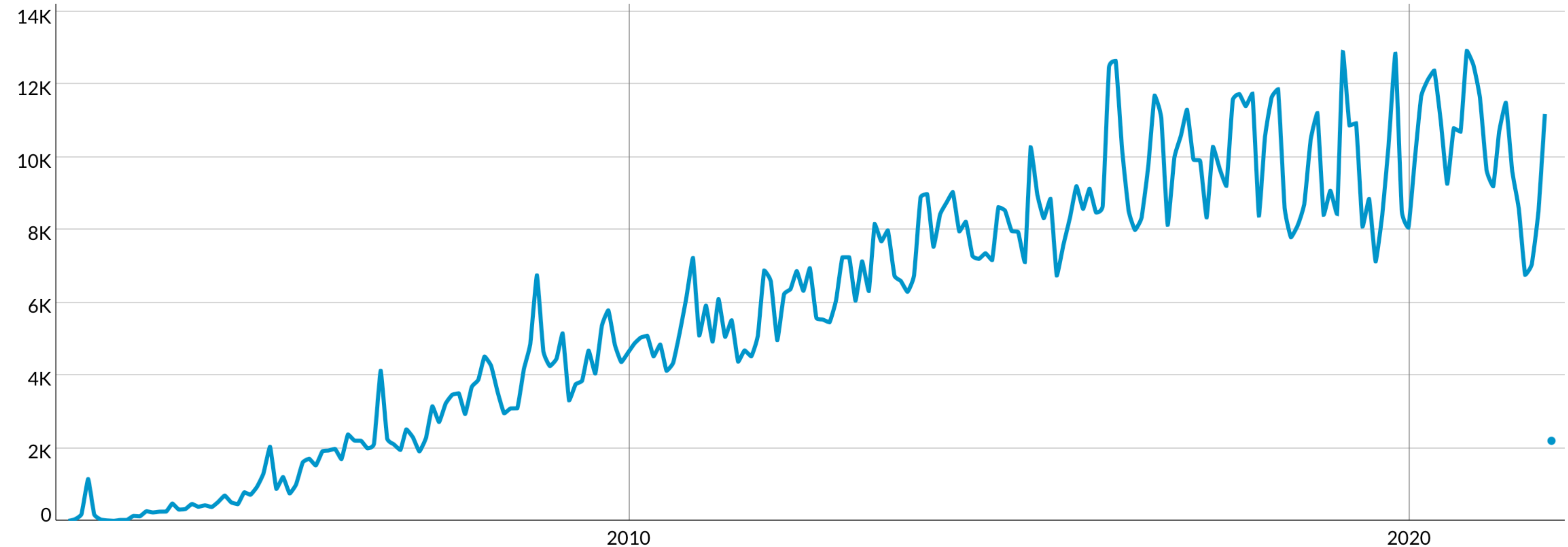
Date Range:

2002-11-28 to 2021-11-07

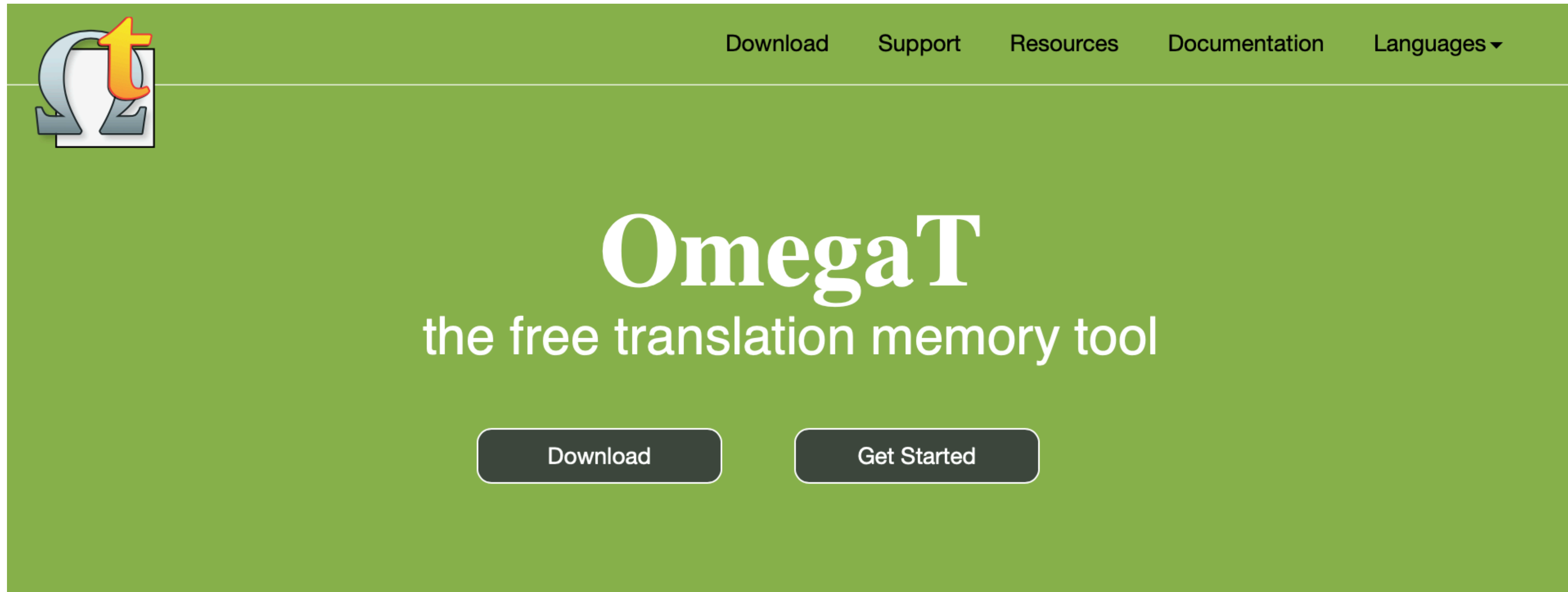
Daily

Weekly

Monthly



We have a cool logo and a cool site too...



OmegaT is a free translation memory application that works on Windows, macOS, Linux... It is a tool intended for professional translators. It does not translate for you! (Software that does this is called "machine translation", and you will have to look elsewhere for it.)

What does OmegaT bring to the game?

- Professional translators have a problem, they get paid for delivering
 - fast,
 - consistent,
 - quality translations...

So they need to use tools that help them to achieve that...

... and sadly, po-mode is not part of the tool set.

Glossary

Free Software Foundation = Free Software Foundation

Software Foundation = Software Foundation

1. (Free Software Foundation), Sun

free software = logiciel libre

1. GNOME

1. Sun

copyright = copyright

1. Sun

software = logiciel

1. Legal term, Sun

1. Sun

free = disponible, libre

1. The state of a volume that is not mapped to initiators. SunGloss : libre, disponible... selon contexte, Sun

2. ATTENTION ! La traduction de cet adj. dépend vraiment du contexte (quelconque, gratuit, disponible, etc.), Sun

Editor – misc/semantic.texi.fr.po

@value{TITLE}

docstyle.texi

docstyle.texi

This manual documents the Semantic library and utilities.

Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation, Inc.

[fuzzy] Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation, Inc.<segment 155152 ¶>

Fuzzy Matches

1. Copyright @copyright{} 1990--1995, 1997, 2001--2021 Free Software Foundation, Inc.

Copyright @copyright{} 1990--1995, 1997, 2001--2021 Free Software Foundation, Inc.

<100/100/86% >

2. Copyright @copyright{} 1990--1995, 1997, 2001--2017 Free Software Foundation, Inc.

Segment Properties

Name	Value
Comment	Extracted comments: type: copying
File	misc/semantic.texi.fr.po
Path	

Comments

Notes

I'm using this segment to show how neat OmegaT is...

Multiple Translations

Dictionary

Machine Translation

Project autosaved on 8:37 PM

LCK | INS 20/286 (6923/150151, 168393) 82/89

Glossary

Free Software Foundation = Free Software Foundation

Software Foundation = Software Foundation
1. (Free Software Foundation), Sun

free software = logiciel libre
1. GNOME
1. Sun

copyright = copyright
1. Sun

software = logiciel
1. Legal term, Sun
1. Sun

free = disponible, libre
1. The state of a volume that is not mapped to initiators. SunGloss : libre, disponible... selon contexte, Sun
2. ATTENTION ! La traduction de cet adj. dépend vraiment du contexte (quelconque, gratuit, disponible, etc.), Sun

Notes

I'm using this segment to show how neat OmegaT is...

Multiple Translations

Dictionary

Machine Translation

Project autosaved on 8:37 PM

OmegaT 5.7.0 :: Projet OmegaT de Documentation Emacs

Editor – misc/semantic.texi.fr.po

@value{TITLE}

docstyle.texi
docstyle.texi

This manual documents the Semantic library and utilities.

Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation, Inc.
[fuzzy] Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation, Inc.<segment 155152 ¶>

Fuzzy Matches

1. Copyright @copyright{} 1990--1995, 1997, 2001--2021 Free Software Foundation, Inc.
Copyright @copyright{} 1990--1995, 1997, 2001--2021 Free Software Foundation, Inc.
<100/100/86% >

2. Copyright @copyright{} 1990--1995, 1997, 2001--2017 Free Software Foundation, Inc.

This part is the Editor

You can see the 3 segments we had at the top of the PO file.

The segment being currently edited is shown with its number.

And a "fuzzy" match has been found in legacy translations and has been automatically inserted.

Glossary

Free Software Foundation = Free Software Foundation

Software Foundation = Software Foundation

1. (Free Software Foundation), Sun

free software = logiciel libre

1. GNOME

1. Sun

copyright = copyright

1. Sun

software = logiciel

1. Legal term, Sun

1. Sun

free = disponible, libre

1. The state of a volume that is not mapped to initiators. SunGloss : libre, disponible... selon contexte, Sun

2. ATTENTION ! La traduction de cet adj. dépend vraiment du contexte (quelconque, gratuit, disponible, etc.), Sun

OmegaT 5.7.0 :: Projet OmegaT de Documentation Emacs

Editor - misc/semantic.texi.fr.po

@value{TITLE}

docstyle.texi

docstyle.texi

This manual documents the Semantic library and utilities.

Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation, Inc.

[fuzzy] Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation, Inc.<segment 155152 ¶>

Fuzzy Matches

1. Copyright @copyright{} 1990--1995, 1997, 2001--2021 Free Software Foundation, Inc.

Copyright @copyright{} 1990--1995, 1997, 2001--2021 Free Software Foundation, Inc.

<100/100/86% >

2. Copyright @copyright{} 1990--1995, 1997, 2001--2017 Free Software Foundation, Inc.

Segment Properties

Name	Value
Comment	Extracted comments:
Type	copying
Path	

Notes

I'm using this segment to show how neat OmegaT is...

Multiple Translations

Dictionary

Machine Translation

Project autosaved on 8:37 PM

That part over there is the Glossary viewer

Matching glossary items are underlined in the current segment

And this popup here

is actually an auto-complete system that not only works with glossary items, but also history predictions, predefined strings, etc.

Glossary

Free Software Foundation = Free Software Foundation

Software Foundation = Software Foundation

1. (Free Software Foundation), Sun

free software = logiciel libre

1. GNOME

1. Sun

copyright = copyright

1. Sun

software = logiciel

1. Legal term, Sun

1. Sun

free = disponible, libre

1. The state of a volume that is not mapped to initiators

SunGloss : libre, disponible... selon contexte, Sun

2. ATTENTION ! La traduction de cet adj. dépend vraiment du contexte (quelconque, gratuit, disponible, etc.), Sun

Editor

@value

docsty

docsty

This manual documents the Semantic library and utilities.

2021 Free Software Foundation, Inc.

2009--2021 Free Software Foundation, Inc.

1. Copyright ©copyright() 1999--1999, 1997, 2001--2021 Free Software Foundation, Inc.

Copyright ©copyright() 1999--1999, 1997, 2001--2021 Free Software Foundation, Inc.

<100/100

2. Copyright ©copyright() 1999--1999, 1997, 2001--2017 Free Software Foundation, Inc.

Documentation Emacs

Segment Properties

Name	Value
Comment	Extracted comments: type: copying References: /Users/suzume/Documents/Repositories/Projet OmegaT de Documentation Emacs - Sources/doc/misc/semantic.texi:30
File Path	misc/semantic.texi.fr.po

Comments

Path:
Comment:
Extracted comments:
type: copying

References:
/Users/suzume/Documents/Repositories/Repositories/Projet OmegaT de Documentation Emacs - Sources/doc/misc/semantic.texi:30

Notes

I'm using this segment to show how neat OmegaT is...

Multiple Translations

Dictionary

Machine Translation

Project autosaved on 8:37 PM

LCK | INS 20/286 (6923/150151, 168393) 82/89

Those parts over there are reference / contextual information that the file itself (po) provides us

Down there we have "Notes" that can be shared between translators.

Below that, we have extra functions that we're not using today: multiple translations for identical segments, dictionary matches, machine translation fetching), along with status info (here, and autosave time)

And down there, we have current project stats that tell us that we have about 143,000 more segments to go...

But we still have about 12,000,000 characters to type...

And

this



```
@value{TITLE}
```

this



```
docstyle.texi
```

and

this



```
Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation,  
Inc.
```

is exactly what we
do *not* want to have
to type...

```
Editor - misc/semantic.texi.fr.po

‘
@value{TITLE}
‘

docstyle.texi
‘

This manual documents the Semantic library and utilities.
<segment 155151 ‘>

‘

Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation,
Inc.

‘
```

Texinfo is not native to OmegaT

- All those things are strings that *must* be reproduced in the translation but are tiresome and error-prone to type.
- OmegaT allows the user to define protected strings with regular expressions so that such strings can be
 - visualized right away in the source segment
 - entered semi-automatically
 - checked for integrity

I love regular expressions...

```
\(?(@code\[^\]*\}+|@command\[^\]*\}+|@kbd\[^\]*\}+|@key\[^\]*\}+|@file\[^\]*\}+|@url\[^\]*\}+|@uref\[^\]*\}+|@email\[^\]*\}+|@\[^\{\s\]*\[^\}\(\)\]*|\)\?\}\)\)?|[CM]-\[^\s\]*[\da-zA-Z-\.]*|^[a-zA-ZX]$|^[a-zA-ZB-Z] |^[a-zA-Z]+(-[a-z]+)+$|(?<=,\s)[a-z-]+(?=\})|\(?(?<=^[^\w][\W\{])[a-z]+(-[a-z]+)+\)?|,\s(?:=[a-z-]+\})|([A-Z]){2,}|(https:|~)((/|\\)[^\];\n\s\"<,[ ]+)|``|' '|,,+
```

And now?

This



@value{TITLE}

this



docstyle.texi

and

this



Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation,
Inc.

<segment 158988 ¶>

are now protected
strings that I can
visualize right away...

and you probably noticed that I cheated a bit because I added
years and the name of the FSF to the earlier regex...

So what?

Undo Last Action⌘ Z

Redo Last Action⌘ Y

Replace with Match or Selection⌘ R

Insert Match or Selection⌘ I

Replace with Machine Translation⌘ M

Replace with Source⇧ ⌘ R

Insert Source⇧ ⌘ I

Insert Missing Source Tags⇧ ⌘ T

Insert Next Missing Tag⌘ T

Export Selection⇧ ⌘ C

Create Glossary Entry...⇧ ⌘ G

Search Project...⌘ F

Search and Replace...⌘ K

Search Dictionaries

Switch Case To>

Select Match>

Insert Unicode Control Character>

Use as Default Translation

Create Alternative Translation

Remove translation

Set empty translation

Register Identical Translation⇧ ⌘ S

Editor – misc/semantic.texi.fr.po

“

@value{TITLE}

“

docstyle.texi

“

This manual documents the Semantic library and utilities.

“

Copyright @copyright{} 1999--2005, 2007, 2009--2021 Free Software Foundation, Inc.

<segment 158988 ¶>

“

@copyright{}| 1999

@copyright{}1999|2005

1999

2005|2007

2005

2007|2009

2007

2009|2021

2009

Missing tags

Press **^+↑** to go to **Character table** suggestions.

Press **^+↓** to return to **Autotext entries** suggestions.

So what? Part II

Copyright @copyright{} 1999--2005, 2007, 2009--
Inc.

1998<segment 158988 ¶>

Issues for semantic.texti.fr.po (4)

Seg...	Type	Description
158988	Tags	Missing x7
158989	Tags	Missing x2
158990	Tags	Missing x2
158991	Tags	Missing x2

Copyright @copyright{}
1999--2005, 2007, 2009--2021
Free Software Foundation, Inc.

1998

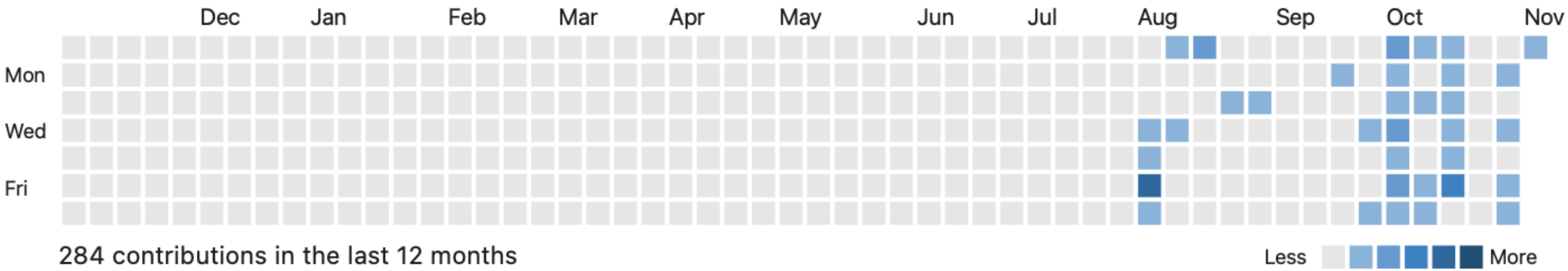
Almost native Texinfo support means

- 248,182 such protected strings \Rightarrow 248,182 semi-automatic entries
- about 2,106,537 characters less to type (-16.7%)...
- about 232,041 words less to type (-12.1%)...
- We're down to 2 years, 3 months, 25 days.
- And if we add the proper regex for `org` we're even better off.

This is still too much work for one translator

- OmegaT is a hidden svn/git client.
- Projects can be hosted on svn/git servers
- Translators only need credentials and OmegaT commits for them
- Translators do *not* have to know anything about VCS
- And nothing has to be done in those ugly and clumsy web-based translation interfaces...

One software to manage them all...



brandelune pushed to [main](#) at [brandelune/documentation_emacs](#)



bb034685ff Translated by suzume

32 seconds ago



Jean-Christophe Helary

800ef1f0d2

Translated by suzume



Jean-Christophe Helary

7c1d69b15c

Translated by suzume



matth

2e8f9bf0a5

Translated by matth



matth

b9e3785473

Translated by matth

OmegaT does all the heavy-duty work

- OmegaT regularly saves to and syncs from the server
 - Translators are kept updated with work from fellow translators
 - OmegaT offers a *not* scary "conflict resolution" dialog when necessary
 - Translators *never* have to do *anything* with svn/git ever...
-
- Imagine, 4-5 translators who translate 500 words a day... Still a lot of work, but it starts to look realistic.

Thank you for being with me today/tonight

I promised I would *not* talk about:

- How to create an OmegaT project
- How to set up an OmegaT team project
- How to use OmegaT from the command line to work in localization pipelines
- How to use machine translation and MT "post-edit"
- How to convert back the translated files to texi format
- How to install translated texi files for use in Emacs

But I'm around, so if you are interested, don't hesitate to ask.