
Easyling wiki docs Documentation

Release latest

December 20, 2016

1	Introduction	3
1.1	Features	4
1.2	Whitelabel	5
1.3	Pricing	5
1.4	Further Information	5
2	Getting started	7
2.1	Registration & Login	7
2.2	Setting up Your First Project	7
2.3	Running a Discovery	10
2.4	Giving a Quote	11
2.5	On the Workbench	12
2.6	The 3-Phase Workflow	12
3	The Dashboard	15
3.1	Introduction	15
3.2	Menu	15
3.3	Cookbook	44
4	The Workbench	51
4.1	Introduction	51
4.2	Moving Around	52
4.3	Views	59
4.4	Translation of Segments	63
4.5	Collaboration	69
4.6	Cookbook	71
5	Maintenance	75
5.1	Project maintenance	75
5.2	Automation possibilities	75
5.3	Automatic pre-translation	76
5.4	Caching And The Proxy	77
6	Technical Reference	79
6.1	Architecture	79
6.2	Request Handling	80
6.3	Classification of Content	81
6.4	Translation Memories	82
6.5	Using TMs	82

6.6	Page modifiers	83
7	Troubleshooting & Support	85
7.1	Contacting Support	85
7.2	Issues	85
8	FAQ	93
8.1	General	93
8.2	Caches	94
9	Glossary	97



Introduction

Easyling is a cloud-based translation proxy solution designed to make websites available in several languages. Easily. A translation proxy is a layer placed between the browser and the original website, through which the visitor sees the original website in a different language.

There are several proxy solutions available on the market, yet Easyling is unique in that:

- it is a solution targeted primarily at LSPs - which also means that this is the **only** LSP-independent solution on the market
- also available as a whitelabel solution
- standard-compliant XLIFF export option ensures CAT-independence
- automated XLIFF export/import option between CAT / Translation management systems and Easyling
- in-context live preview available in certain CAT-tools at translation-time

What does that mean?

If you are a **business owner**, it can help you reach a wider customer-base through providing information to your potential customers in their native language. What's more, you can do more than just translating the text on your website, you can **localize** it: you can also adapt your message, the images displayed, or even your product range offered to the targeted culture. And all this without the need of heavy upfront investment in IT infrastructure and personnel, and the hassle with regular maintenance and upgrade. Easyling takes care of the IT part, so that you could concentrate on the content - and growing your business.

If you are a **language service provider** (LSP), you can offer cutting-edge website localization services to your customers - even under your own brand name! Easyling provides the technology, takes care of the IT infrastructure, leaving you to concentrate on your core business: cross-cultural communication. What's more, your translators don't need to learn using just another tool, they can keep using their own preferred CAT-tools.

Sounds good?

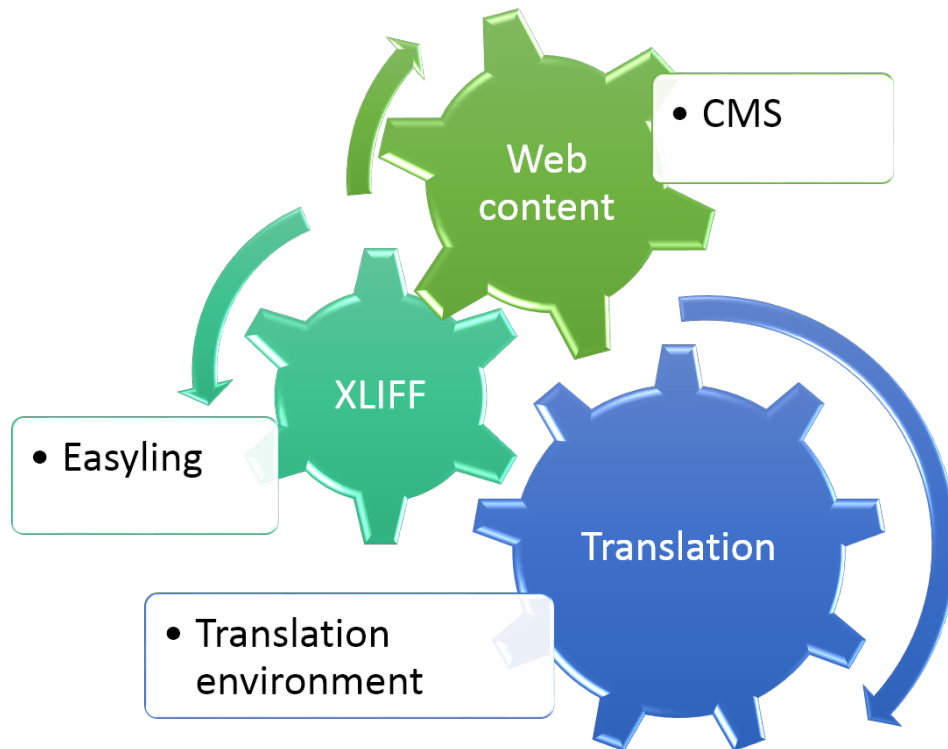
There are several challenges both business owners and language service providers face during website translation. The "ideal" workflow would be to create the content in the original language, get it translated into the desired languages, and then publish all language variants at the same time, from the website owner's own content management system (CMS) - right from the very first page on the website. But reality is different. Apart from the fact that not many CMSs are capable of handling several languages, usually website localization comes into the picture at a later stage, when there is already a huge amount of data published on the website. And, in most cases, the website owner can't extract the content for translation. If they can't extract the original, there's no easy way to load the translated content back either. Furthermore, if the website owner can't extract the content into translatable format, it is impossible to get a proper estimate for the translation costs in time and money...

Easyling can, however, discover the website by following links and grabbing translatable and localizable content - and convert it into a translatable format. This gives a realistic view of the magnitude of the translation task, and, thanks to the translation proxy, even a partially translated site can give full user experience on the website visitor's side.

Data can be extracted with a couple of clicks - and the publication of the translated site is similarly easy.

Where do we fit in the localization workflow

Easyling bridges the gap between the CMS and the translation workflow; it enables you to extract content from the website in a translation-ready format that can be used in *any* translation environment.



- If you are a content owner, it means that you have a technology solution that enables you to choose any LSP that suits your language requirements the best.
- If you are an LSP, it means that you can take care of the website translation requirements of your clients.

In either case, technology will no longer be the bottleneck.

1.1 Features

- process HTML, JavaScript/AJAX/JSON, XML (note: translation of Flash is not supported)
- Use the X-Proxy and other Preview mode domains to see everything in context. use Advanced Settings to translate text coming from JSON/XML sources.
- Automatically crawl static pages / HTML content only. Add extra AJAX URLs with the proper parameters.
- Fine-tune your settings to help the crawler decide what URLs to handle as same, and what to visit looking for new content.
- Translate forms, messages and dynamic content
- Translate images: replace them with their localized counterparts on a target-language basis.

- Link any external domain via multiple projects.
- Modify page behavior using customized CSS and JavaScript injection.
- Use regular expressions to filter for content.

and many more!

1.2 Whitelabel

Easyling offers a white label version that can be customized with your corporate logos and domains to create a branded version, allowing you to use & sell Easyling as your own product. In order to create the branded version, seven criteria must be met. See the “Whitelabel setup” section for details.

1.3 Pricing

Our pricing follows the ‘pay-as-you-go’ model, so you only get charged for what you use. The total cost is made up 2 types of fees: one-time fees and a monthly charge.

1.3.1 One-Time Fees

- **Discovery:** 1 EUR or 1.2 USD / 1000 pages for every Discovery
- **Scan:** 2 EUR or 2.4 USD / 1000 source words for each **new** word during a Scan (if no new words are found, a Scan counts as a Discovery)
- **Translation Memory:** storing your (human or machine) translation in the database costs 10 EUR or 12 USD / 1000 source words / target language

Content extraction and translation memory fees apply only the first time around, which means 102% repetitions are not counted. Once a segment is stored, subsequent scans will treat it as repetition and **no additional charges** will apply.

1.3.2 Monthly Fees

A 1 EUR or 1.2 USD / 1000 page views monthly fee applies to serving your translated content over the proxy.

In exchange, you get a guaranteed 99.99% website availability, and a capacity to handle practically unlimited traffic. You also have the option to serve the translated site from your server (in which case no proxy fee will apply), but in this case availability and traffic handling depends on your infrastructure.

1.4 Further Information

Visit us at <https://easyling.com> and sign up for a dedicated demo if you are new to translation over the proxy!

See <https://easyling.com/pricing> for pricing details. Try our Price Calculator, too!

Contact Support: support@easyling.com

Be sure to check out all the other tutorials we’ve compiled:

Tutorials

- <https://www.youtube.com/watch?v=S47kArNiJ1o>
- <https://www.youtube.com/watch?v=8VsBy2bGo64>
- <https://gitlab.com/easyling/wikis/home>
- <https://drive.google.com/open?id=0Bw53oZELMrf8V1FIUnhmNEtubTA>
- <http://lesson101.tutorial.easyling.com/>
- <http://lesson102.tutorial.easyling.com/>
- <http://lesson103.tutorial.easyling.com/>
- <http://lesson105.tutorial.easyling.com/>

Release notes

- <https://www.easyling.com/category/blogpost/release-notes/>
- <https://www.easyling.com/change-logs/>

Getting started

Let us give you a quick overview of how the proxy is used. In this section, we give you a quick introduction to the Dashboard and we'll also set up a simple project.

2.1 Registration & Login


To use Easyling, you need to register and set up an account for the service at <https://app.easyling.com>. You can start using the service right away after registration.

Logging in, you will be taken to the Dashboard, which is the project management center, every detail of which we'll get to in later sections of this manual. There will be no existing projects at the outset, so let's try setting one up.

2.2 Setting up Your First Project

To do this, click on **Create new project** dropdown at the top and choose **Add project**.

This opens the **Add project** dialog box, where you can enter the URL of the website you would like to translate, and also select the website language; this sets the source language of the translation project. Click on Advanced Settings to access extended functionality, such as custom SRX files.



Add project

Website URL

You can enter full path as well,
as starting page to discover the site

Website language

☒ **Start discovering automatically**

☐ **Process pages in source language only**
(requires [Google language detection API](#))

..... ▲ Advanced options

☒ **Check if the domain redirects to another domain**
E.g. <http://example.com/> redirects to <https://www.example.com/>

Include pages only starting with

/our-company
/portfolio
/contact

Leave it empty to include all

Ignore these paths

/forum
/blog
/news

Enter unwanted paths (Forums, blogs, etc.) here to prevent them from
showing up in Pages.

Provide custom SRX file (optional)

No file chosen

SRX stands for Segmentation Rules eXchange.
Text will be broken into sentences according to these rules.

2.2.1 Add a Target Language

You will also need to add your target language(s), so use the option on the Dashboard to add then to the project. It's not just that there is not much to do in terms of translation without a target: many crucial features (including the Preview proxy and the entire Workbench) are entirely unavailable as long as no target languages are set.



You can add any number of target languages. You can use the search function to lookup languages based on locale code or country name.

2.3 Running a Discovery

Discovery limit

Discovery lets you map out the site structure before the actual content extraction. Running it consumes one pageview for every visited URL from your balance, but does not store content.

Let's discover pages max (leave blank for Unlimited).

Page freeze and dictionary freeze are enabled.

☐ Add & process newly found pages (takes precedence over Page Freeze setting!)

☒ Add & process newly found resources

☐ Add & process newly found image resources

☒ Skip content type check if already determined (improves re-scan time, safe to use)

☐ Limit crawl depth to

☐ Ignore pages matching the following regular expression during this crawl

☐ Limit number of simultaneous requests (1-8)

☒ Use current exclusion rules ☐ Use temporary exclusion rules

☐ Build source cache ☒ Preserve & use existing source cache

Select cache:

[Source cache settings](#)

The next step is to figure out what (and how much of it) to translate and do it from a single target URL. For this, a Discovery has to be run on the site.

Discovering a website means running a `crawler` on it and allowing the proxy to ‘explore’ it in its entirety in order to provide Statistics for you. As you can see in the ‘Add project’ dialog window, the Dashboard is set up to automatically start a Discovery on a webpage after creating a project - but don’t worry! After clicking on the **Add project** button, a new dialog will open where you get to set up additional details of the Discovery before really starting the process.

There are many details that potentially have to be taken into account when setting up a truly effective Discovery, but let’s set those aside for the moment. For now, simply Click on the ‘Add project’ button and after the Discovery dialog opens, click on Discover to start a crawl on the website. **The default settings are safe.**

Depending on the size of the site, a Discovery can take quite some time to finish. A spinner on the Discovery page will indicate that the system is currently working, but there is a default page limit of 100, which means that if the Discovery finds more than one hundred pages, it will automatically exit, allowing you to fine-tune your settings.

After the process is over, you’ll receive an e-mail about the results. You also have your first Statistics from the Discovery - a word count total from all Discovered pages.

2.4 Giving a Quote

You can use the results of the Discovery to give a quote (based on **unique** source words) to your clients about the estimated work-hours and expenses you forecast for a given project.

The results are an accurate indication of the translation costs associated with the text. However, with websites, it is prudent to **consider other (technical) details** before taking the word count results of the Discovery process at face value.

Investigate the source site and consider the following:

- Is there a great deal of **dynamic content**?
- Any **Site Search** functionality? A **webshop**? A **forum**?
- Any other **web apps** that would have to be localized?
- Do the **average word lengths** of the source and target languages differ significantly?
- Is the **direction of the target language** different than that of the source language?
- Which pages are targeted for translation? Which pages need to be excluded? Ask the client if they have a specific page list.
- does the site have mixed-language content? If yes, ask the client to specify the source language(s) they need translated.
- is there an extant Translation Memory that could be used?
- is there any region-specific content? Does the site use geolocation?
- is there any content behind a secure login?
- are there any subdomains? example.com and blog.example.com require two separate projects that need to be linked.
- are there any other special requirements?
- is there any Javascript-generated content?

If you answered yes to any of those questions, that will require some deliberation, often beyond the primary focus of translators: UI fixes and a measure of fiddling with what's under the hood - take those expenses into account when you make your quote.

NOTE!: If you are unsure as to how to go about translating a part of a website, feel free to contact our **Support Centre** and we'll help you get an accurate picture of the required effort and costs.

It is also advised to negotiate the expected workflow with the client at the quoting phase. The translation of a website is, in most cases, a never-ending process, as new content is added to the original site at certain intervals.

The question is, how content added *after* the initial quote should be treated - both from a technological and fiscal viewpoint. It is a good idea to ask the client about their intentions for update cycles and new content.

Do they wish to publish at the same time in all languages? Or publish on the original site without delay, and then publish the translations later, as they become available? The different options will require different approaches when you get to the maintenance phase of the project.

As a translation proxy is practically a translation layer on top of the original site, serving translations from the datastore by replacing original content on-the-fly, new content will not be replaced, as translation is not stored for that. In practice it means that newly added content will appear in the original language on the translated site. This is called *bleedthrough*. There are 2 approaches to this phenomenon: let bleedthrough happen, to make new content available right away, even if it is in a different language, or block new content from appearing until translation is done. Both have their clear advantages and drawbacks, so you have to discuss with your client which option is more acceptable for them - and set up your project accordingly.

2.4.1 Sales tool for mass production

Easyling also offers a Sales Tool to help LSPs and freelancers in growing their business.

If you have a well-defined group of potential customers you'd like to offer your translation services to, like hotels or restaurants with only monolingual websites in your area, easyling makes it easy for you to impress the business owners. Just collect the URL addresses, add them to the Sales Tool, and easyling will automatically create a project for all webpages according to the settings you specify. Once the translation and post-editing of the translated main pages are ready, you can send a link to the business owners. If your potential clients are impressed with the translated page and the fact that no IT involvement is required on their end, you have a better chance to win the deal.

For more information see our [blog](#) or this short [video](#)

2.5 On the Workbench

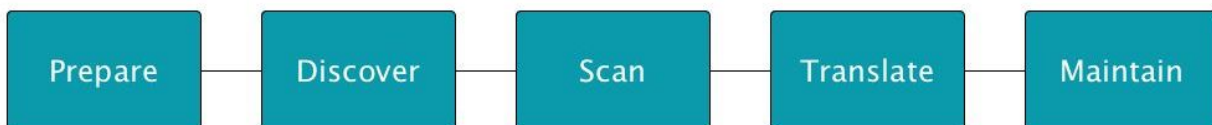
You can export all source segments, translate them in your CAT tool of choice and then reimport your results. But going through that cycle for every small change would get rapidly tedious - wouldn't it be great if you could edit & control your translations in the cloud, where it would all update in real time? You're looking for the Workbench.

In Pages View, you can hover over any page - a menu will show up right next to it - choose 'Translation in List View' and you'll be taken to the Workbench in a different tab.

If the Dashboard is the Project Management Center in Easyling, then the Workbench is the cloud CAT tool, where translation itself takes place. There are many features you can use in the Workbench to make working with websites easier - see the 'Workbench' section of this manual for the details.

2.6 The 3-Phase Workflow

Barring some detail (withheld for the sake of a convenient introduction) the above process is all that you need to get a website translation project going.



Our idea of a project's lifetime can be summarized in the 3-phase Workflow.

2.6.1 1. Discover & Quote

Set up a project and Discover it. Have a Unique Word Count total and a general idea of any technical issues involved. Give your quote to the client (perhaps demo/impress them via the Live Preview). Win the bid.

2.6.2 2. Ingest & Translate

After you are entrusted with the project, collect all text content into a database (overcoming any technical issues that may arise in the process). When you have your data, export it to your CAT tool of choice or translate in the Workbench to a selection of target languages. Reimport and edit. Use our Proofreading and Workflow features to ensure quality.

2.6.3 3. Publish & Maintain

After the translation is greenlit by the proofreaders, you can verify the serving domain and publish the translated website. Add a language selector to the source site. Generally, it is with publishing a website that a deadline is met.

But don't forget that a website is a living thing, with new content arriving every day - the final stage of website localization is always maintenance - making sure that new content gets translated according to schedule, all the while ensuring that visitors to the site will not be inconvenienced by bleedthrough of untranslated content.

Maintenance is the "long tail" of website translation - there are a variety of features in the proxy that make it a lot easier than it would be anyway else.

In the following pages, You will find everything there is to know about using the proxy. Keep reading!

The Dashboard

3.1 Introduction

The Dashboard is your command center. It contains a variety of features you can use to manage your projects. In this manual, we'll take these options in the order that they appear in the menu on the left side of the screen.

3.2 Menu

3.2.1 Dashboard

Project statistics

Please see the “Statistics” section in Discovery for a detailed explanation of Project statistics.

Alarm

Alarm

Send out email notification when any of the alarm conditions is triggered.

Snooze for minutes between notifications.

☐ Trigger alarm above page view / second limit.

☐ Trigger alarm above words added to memory / second limit.

☐ Trigger alarm above words translated in memory / second limit.

You can set up alarms to be notified of any actions in your project that could potentially lead to undue expenses. Enter the value for the timeout (snooze) between subsequent e-mail notifications, to prevent the Alarm from flooding you with e-mails. After that, check one of the alarm options and enter the limit above which you want to receive a notification. You can set an alarm for:

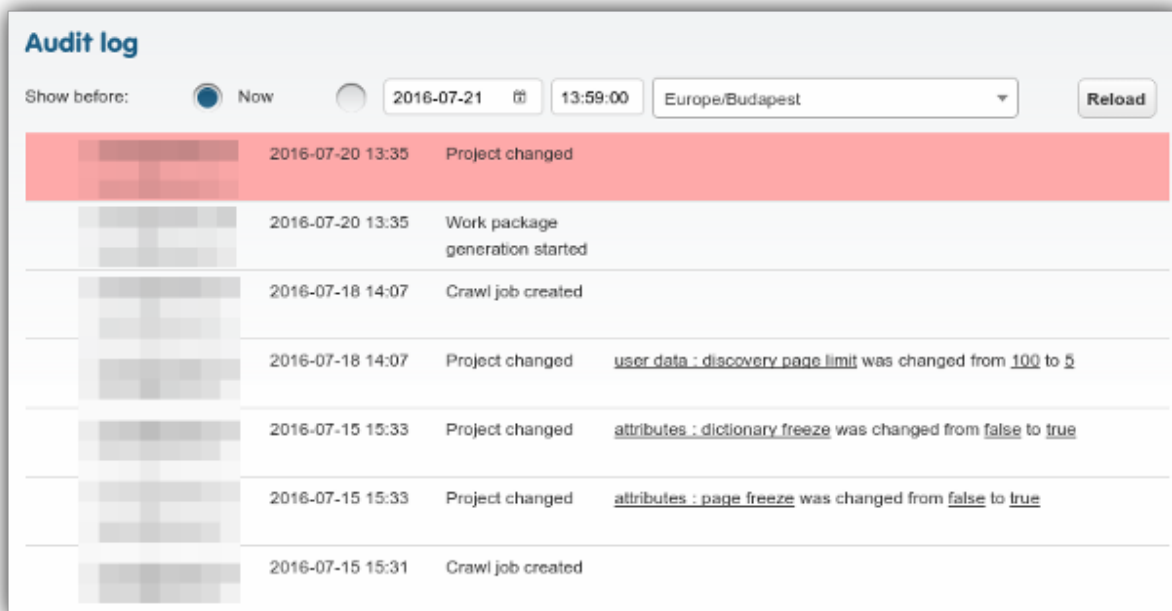
Page Views: if you wish to detect unusually large amounts of requests to the proxy. As Scans and Discoveries count as Page Views, you can also use this alarm to detect if a Scan was started.

Words added: if you wish to receive notification about new content that exceeds the set amount.

Words translated: if you wish to receive notification about Machine Translation or reimport events on the project.

The e-mails are sent to the owner by default, but you can add the ‘Receive notification emails’ right to any participant in Dashboard > Sharing settings.

Audit log



Audit log		
Show before:	<input checked="" type="radio"/> Now	<input type="radio"/> 2016-07-21 13:59:00 Europe/Budapest <input type="button" value="Reload"/>
	2016-07-20 13:35	Project changed
	2016-07-20 13:35	Work package generation started
	2016-07-18 14:07	Crawl job created
	2016-07-18 14:07	Project changed user data : discovery page limit was changed from 100 to 5
	2016-07-15 15:33	Project changed attributes : dictionary freeze was changed from false to true
	2016-07-15 15:33	Project changed attributes : page freeze was changed from false to true
	2016-07-15 15:31	Crawl job created

The audit log is the minutes-taking feature of Easyling. It records all activity on the project with a username/e-mail address attached to each entry.

Some Audit Log entries can get quite long, such as when Page Modifier or CSS overrides are being edited and saved multiple times. For this reason, the Audit log displays only the first three lines of long entries by default. Use the grey bar with the downward-pointing arrow under a log entry to see the full text of that entry.

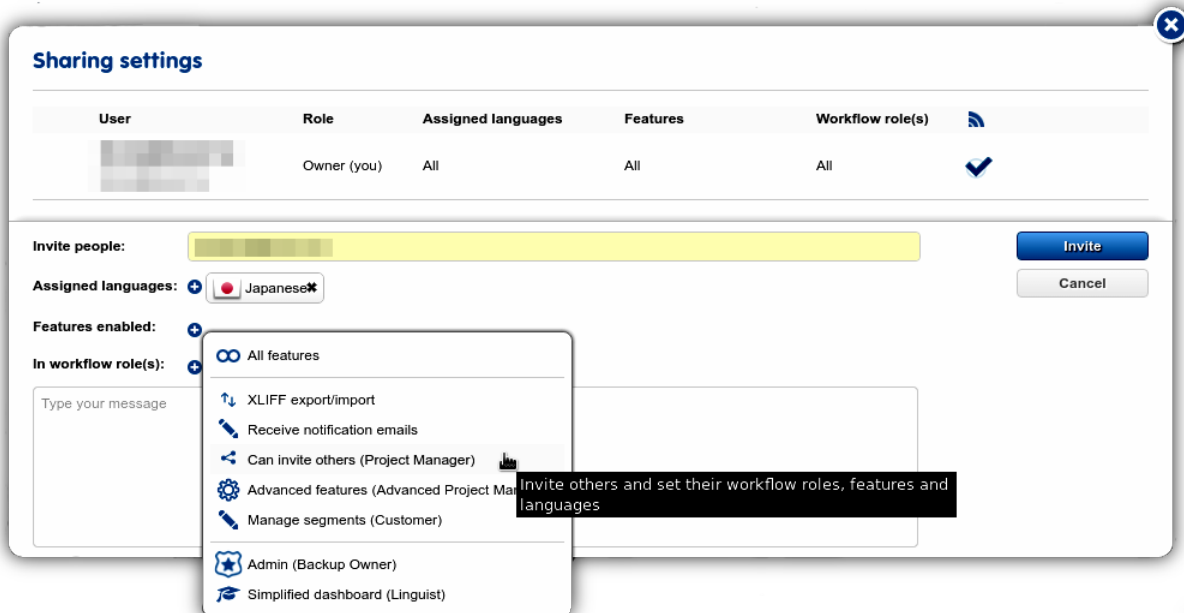
Sharing settings

In the world of website localization, it is rare for a project to be realized by a single person. As project owner or backup owner, use the Sharing settings dialog to invite people into the project and assign roles/editing roles to any one of them.

At the bottom of the Dialog, you will see the line, “Invite People:” if you click on the field, the invitation settings part of the dialog will slide up, allowing you to enter further details of the project participant to be added. You can assign languages, enable admin/access features and workflow roles for each user you invite. You can do this by clicking on

the blue plus icon next to each of those options. Make your choices in the small pop-up lists. **These settings are not final, you can change them later.**

Features



The various **features**, (administrative rights) deserve to be detailed, Here they are:

All features: every feature you see below the separator in the list will be added for the user. Note that this has a potential to cause problems: If, for example, you simultaneously set **Advanced features** and **Manage segments**, it results in an editing rights conflict. There are security implications you need to consider when giving your users rights, so use this feature with care.

XLIFF export/import: enables XLIFF export/import of segments for the user. Especially important for those users who will be coordinating the translation effort throughout the project.

Receive notification e-mails: Add user to the list of recipients of Discovery/Content Scan/Alarm, etc notifications.

Can invite others: designates the user as a Project Manager, able to invite others onto the project. Take care not to include other, conflicting roles, such as **Customer**, which could re-restrict access.

Advanced features: give the user the power to edit languages as well as any entry in the project, and editing most setting, up to, and including, the URL inclusion-exclusion rules. Segmentation, publishing and certain advanced settings remain beyond the reach of the user.

Manage segments: make the user a Customer, with the ability to manage (approve or exclude) pending source segments. The user will also require a workflow role and an assigned target language to access the Workbench to manage segments. **WARNING! This right restricts the user to this role, other features will be disabled!**

NOTE Features are predominantly important from a project management viewpoint. **Workflow roles**, on the other hand, are used to manage user interaction **on the Workbench**, based on expertise (whether a user is a translator, proofreader or client will decide what sets of segments are available to them for editing - see the documentation of the Workbench for details).

After a user redeems their invitation using the link they receive in an e-mail, their address and username will appear under the owner. Their features and workflow roles can be edited in the list.

You can invite as many participants to the translation project as you'd like. You can also enable e-mail notifications for any of them by checking the notification icon in the last column.

Project Roles

Based on the project features that are available for a user, the following project roles are available:

- **Owner:** Every project has an owner with unlimited powers over the project: the owner may add or remove anyone on the project, edit any entry in any language, including adding new languages, any change any setting, including the advanced ones. There can only be one owner on a project, but owners may renounce ownership, designating another user and setting their own privileges.
- **Backup Owners:** same as project owners. They can change any setting and entry, add or remove users, and modify language settings.
- **Project Managers:** can invite others onto the project. Other features and roles can be added as well. Take care not to include other, conflicting roles (restrictions override rights)
- **Advanced Project Managers:** they can edit languages, as well as any entry in the project. They can change most setting, up to and including the URL inclusion-exclusion rules. They cannot change segmentation rules or publishing settings. A few Advanced settings are also beyond their reach.
- **Linguists:** have access to the Simplified Dashboard and Content menu. They are only allowed to edit segments in their designated language and workflow role. They can receive notification emails about project updates, and may be given the power to import/export XLIFF files.
- **Contributors:** default users, capable of editing any entry in their selected language and workflow role, but nothing else. They may receive notification emails and project update emails, but they may not edit their features, nor invite anyone else, nor access any of the advanced settings.
- **Customer:** This role intended for a representative of the customer who exercises oversight over content. This role is read-only, meaning despite the target language added to the user, they are unable to actually make edits. They can *Approve* or *Exclude* segments.

Client Approval Sometimes there is need for the client to view and check new segments before they enter the translation workflow.

In the **Advanced Settings** menu, default state of new segments can be set to the following:

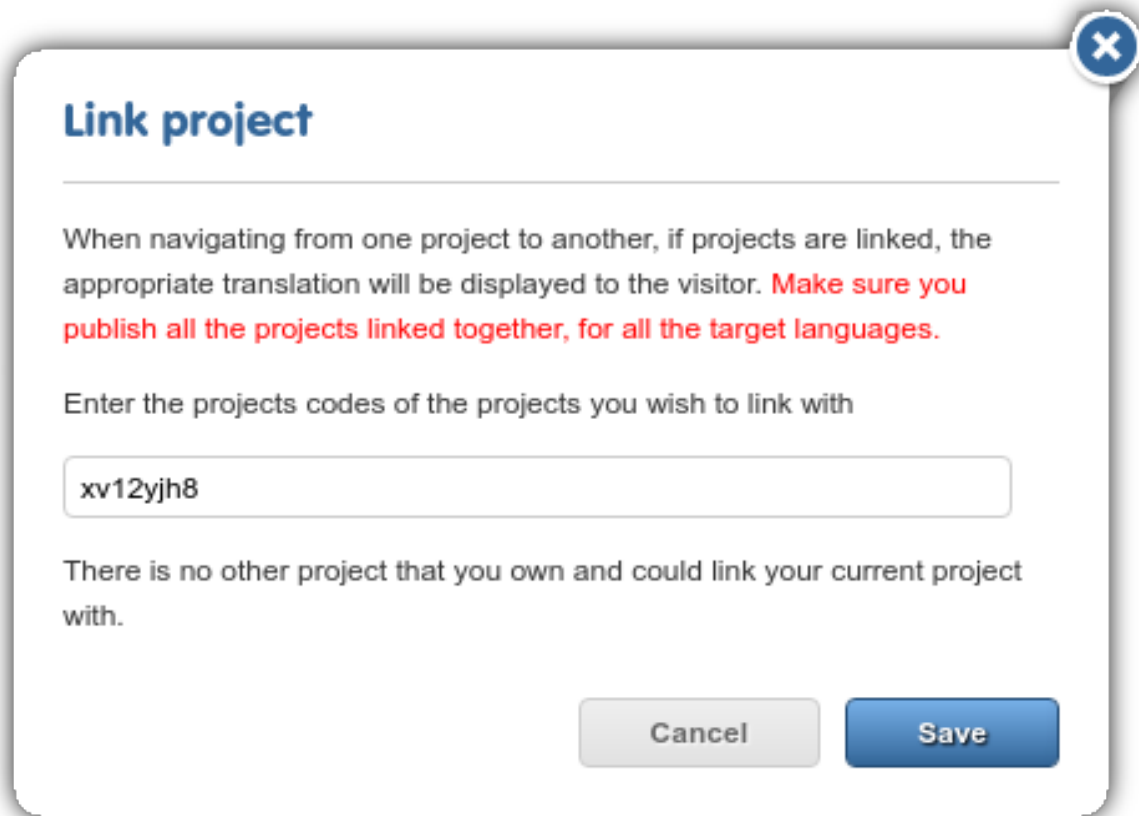
- Pending
- Excluded
- Approved

When set to Pending or Excluded, new segments picked up will acquire that state automatically and manual approval is needed. In the Sharing Settings menu, the customer role can be assigned for this purpose.

Important: The client role is read-only. Anyone assigned this role won't be able to edit content, but approve them for translation. If you are unable to edit segments on the Workbench, please check Sharing Settings first.

Link projects

This dialog has a single input field for project codes and looks like this:



Use it to link two projects together. For example, **example.com** and **blog.example.com** are separate domains, so two separate Projects (and Scans!) are required to create a full translation for the site. When both projects are ready to be published, use this function to link and serve them together on the proxy.

Page Caches

Source Caches

SCs are your first line of defense against bleedthrough: it is a way of storing and reusing content that was received and picked up from the **remote server** (in other words, the **source site**). The goal of Source Caches is to create a *controlled snapshot* of the source content that you can safely use to translate and serve content over the proxy.

You might want to set up one Source Cache for publishing content and another to be used during translation. The content of the Source Cache being published should be highly translated, and new content should be collected in a different Source Cache to be used in Preview mode and during translation.

Benefits of Source Caches

- If you have at least one Source Cache enabled and built, you have essentially decoupled new content from what you are serving over the proxy (you are in control of untranslated content).

Note: Source Caches can only be built by Scans. They will *never* pick up new content unless you expressly run a Content Extraction to update them.

See Case Study #1 for an in-depth coverage of this scenario.

- Multiple Source Caches let you decouple published content from untranslated material - you are free to update any Source Caches you are not currently using for published content over the proxy. This way, you are always serving 100% translated content (using the Cache that is highly translated), and you can schedule for translation updates (on all the others you are free to update anytime with a Scan).

See Case Study #2 for an in-depth coverage of this scenario.

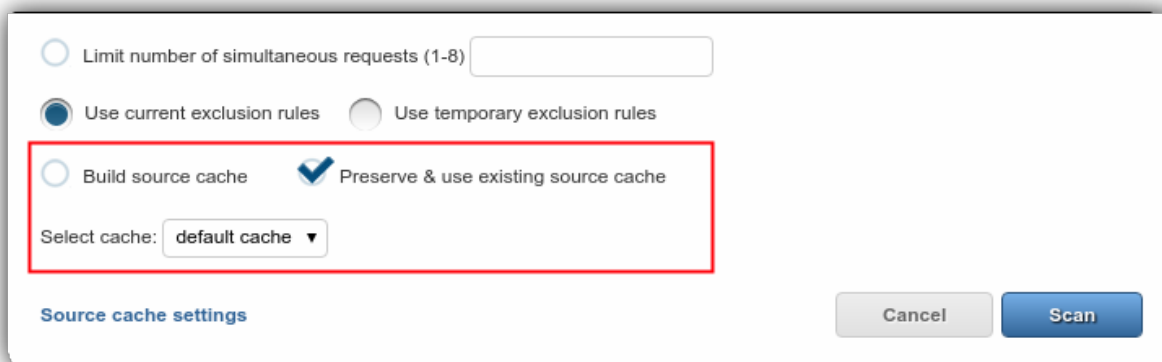
When to build? The best time to build a Source Cache for the first time is during the first Content Extraction. By this time, you should have used Discovery a couple of times to gain a general understanding of the site structure and word count, and you have all necessary exclusion rules set up.

When you are confident in your Content Settings, set up a Source Cache before you run Content Scan (see below for setup directions) on the webpage and build it. This way, you will have a Source Cache that contains all content you're basing your quote on.

Setting up a Source Cache You can enable and create a new Source Cache by checking **Enabled** in the *Page cache* menu in the Dashboard. Note that at this point, the SC is enabled but **unbuilt**.

WARNING!: You have to build the Source Cache before it becomes functional!

Building the SC After enabling the Source Cache, you need to Scan for content to build it. In the Content Scan Dialog, some options become available:



The screenshot shows a dialog box for 'Content Scan' with several options. A red rectangle highlights the 'Build source cache' and 'Preserve & use existing source cache' options, along with a 'Select cache:' dropdown menu. The 'Build source cache' option is currently unselected, while 'Preserve & use existing source cache' is selected with a blue checkmark. The 'Select cache:' dropdown is set to 'default cache'. Below the highlighted area, there is a link for 'Source cache settings'. At the bottom right of the dialog are 'Cancel' and 'Scan' buttons.

Check **Build source cache** and run the Scan. As a result of building the SC, all content that was set to be picked up by the Content Scan is added to the current SC.

If you are updating a SC, you can use the dropdown menu to select the cache you'd like to overwrite.

The '**Preserve & use existing source cache**' option ensures that a SC will not be written to during a Scan. Having both '**Build source cache**' and '**Preserve & use existing source cache**' enabled for a Scan will currently result in an inconsequential action (except for pages that are not in the current cache but referred - these will then be loaded into the cache).

Custom settings If you have more than one Source Caches that you've built throughout multiple Content Scans, you can visit the Page Cache menu again and configure which Source Cache should be used for which proxy mode, as shown in the screenshot below:

The cache is not overwritten if the content served matches the content received (i.e. no processing was done on it), nor are entities larger than the hard-coded maximum entity size (960kb) saved. Furthermore, if a site changes its contents too fast (there are too many cache misses, i.e. the cached content differs from the actual), the Proxy will stop caching the given entity to prevent overusing the database. Should this happen, caches must be cleared manually to restore normal operation and reset the cache miss limit.

What is it for? The purpose of the Target Cache is to reduce the number of requests to the **source site** and decrease response times on the proxy by skipping the content replacement process (which is basically the entire document pipeline) if the content on the **source site** or - if enabled - in the Source Cache remains *unchanged*.

Enabling Target Cache Go into the **Page Cache** menu in the Dashboard and enable Target Cache in the dialog that opens. You can also add a maximum of **5** custom-named Target Caches.

Building the Target Cache After you enable it, the Target Cache is filled exclusively by Page visits on the pretty (published) domain. Previews/visits to the temporary domain will have absolutely no bearing on the contents of the Target Cache.

While separate from the Source Cache, the contents of the Target Cache will nevertheless be determined by your Source Cache settings. If you have no Source Cache, then the Target Cache will be filled by whatever is returned from the source site for a request (a page visit). If you have a Source Cache, then the Target Cache will contain whatever the currently enabled Source Cache is making available to the visitor on the proxy.

The ‘Keep Cache Strategy’ The Keep cache is actually a serving mode of the Target cache, used mainly to prevent the source language bleed-through effect, or Bleeding Effect in short. It operates by checking the cached content’s translation ratio against the remote server’s response. If the cached content is found to have a more complete translation (i.e. its translation ratio is higher), the remote response is discarded and the cached content is served instead. This does not mean an increase in page load speed, but by preventing yet-untranslated elements from appearing in the served page, the Bleeding Effect is eliminated entirely. As new translations are entered into the database, either manually or via XLIFF imports, the difference between the cached response and the actual response decreases, and the newly-translated elements are displayed automatically. Additionally, this check is run every time the source content is found.

Case Studies

Case Study 1: Freeze Site Before Translation begins In this example, the client has declined to use a staging server, as well as declining to halt content updates for the duration of the initial translation. They do, however, insist that bleedthrough must not occur at any point in the translation process. This results in a continuously changing source that makes it nearly impossible to achieve 100% translation.

In this case, the solution is to enable the Source cache, and populate it with a Scan before the first round of translation commences. This creates what is effectively a static snapshot of the site, which remains the same regardless of the updates the client makes in the meantime, providing a stable environment for the translation and review processes.

By using a Source cache to serve the translated sites, they are decoupled from the original, and content updates there will not be reflected in the translations.

However, source content will accumulate in the meantime, and once the cache is purged, bleedthrough will occur! To counter this, the second scenario can be enacted.

Case Study 2: Decoupling Content Update from Ingestion and Publishing For this example, the client has declined to use a staging server to allow you to ingest the content ahead of publishing for translation. They did, however, agree to notify you once new content is published, and have acknowledged that this will cause the translated sites to lag behind the original until the translations (and possibly reviews) are completed. Alternatively, coupling into

the previous scenario, the initial translations are in place, but the source site has moved ahead in the meantime, with the translated site being served by the initial Source cache.

Once the initial translations are completed, the caches are set up according to the image below:

Page cache ^[?]

Source cache

☐ Disabled
☐ Enabled for all proxy modes
☒ Custom settings

Published site: 2016-02-10 ▼
 Live test mode: 2016-07-29 ▼
 Preview mode (-p): 2016-07-29 ▼
 Highlight/popout view: 2016-07-29 ▼
 All other modes (-x, etc): 2016-07-29 ▼

Target cache

☒ Disabled
☐ Enabled

Source caches

#1	2016-02-10		
#2	2016-07-29		

Add new source cache

Target caches

#0	default cache		
----	---------------	--	--

Add new target cache

By driving the published site from a separate cache entity, you gain the ability to decouple the content ingestion cycle from the translation and update cycles.

From this point on, the published site will not reflect any updates until the assigned entity is refreshed. By specifying the update cache in the Scan Limit dialog, you write the server responses into the newly created cache, which will drive the Preview and Highlight modes, allowing you to conduct the translations and ICR using the new content. Once the client signs off on the translations, you simply switch the Source cache entities being used (and purge the Target cache, if in use).

At this point, the previously-live cache entity is freed up for purging and re-building in the next update cycle. Thus, at the next crawl, the default would be selected to receive the updates, leaving the other entity untouched.

By managing which entity is being used to serve the published site and which one is being written to, it becomes

possible to replicate the behavior of the staging server, albeit at the cost of increased attention to detail.

Path settings - Cache & Cookie, Content Type Overrides

Here, you can add target domain paths and define Cache and Cookie overrides for them. Enter a new path (you can set it to be a prefix or an exact URL), then click on Add Path. A New rule dropdown is displayed. You can choose either Cache and Cookie Overrides, Content Type Overrides, or both. After selecting one, the dropdown will disappear, but you'll still be able to add the other using the plus icon that appears next to the path.

By clicking on 'Edit rule', you can set the Cache override to public, private or ignore according to your needs. set the maxAge and override type (whether to clear or use cookies) and you're set.

Use the Save icon next to the rule name or the blue 'Save all' button to save your settings.

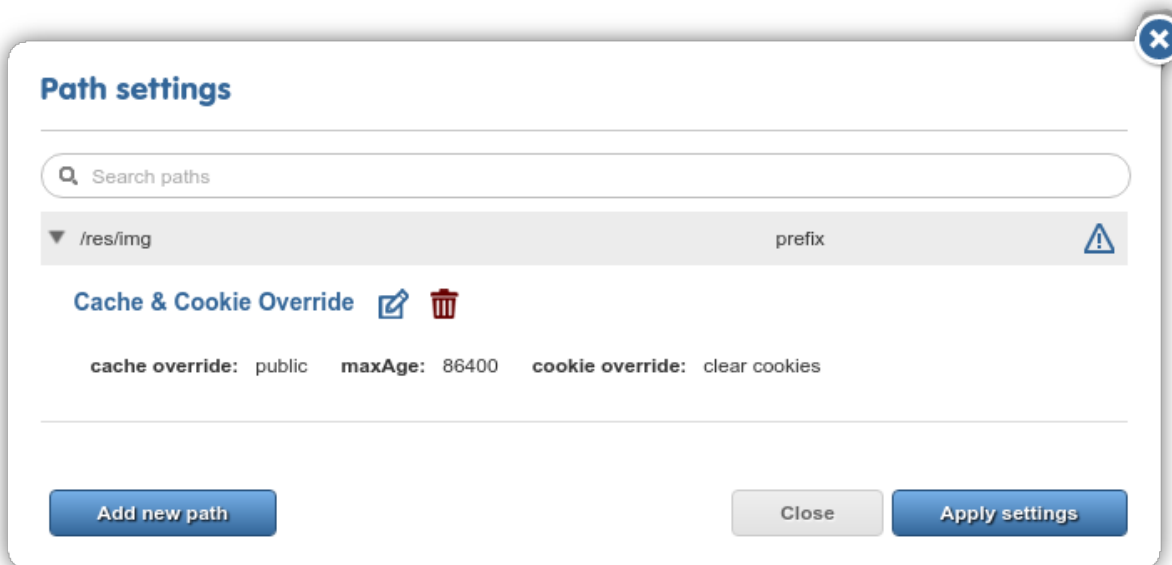
You can always return here and add new paths and rules using the 'Add new path' button if there is a need for it.

Cache Header Overrides

Loading a page in the browser usually involves multiple requests for resources on the original site and other domains. A page can use multiple static resources (such as images) that require no localization.

If an image is referenced by relative name, such as `/logo.jp`, the hand of the proxy is forced: it has to assume the domain to be the proxy domain. The access path of the resource will become `fr.example.com/logo.jp` assuming a French translation of `example.com`. This is a page request over the proxy, and therefore costs money.

Another common issue that lead to this feature is that static resources often have suboptimal Cache Headers. This leads to the same, unchanging content being served from the original server for each request individually, potentially decreasing the efficiency of the site and certainly adding to the number of requests that would have to go through the proxy. The Path Settings dialog can be used to override the default Cache Headers of resources of a given path or exact URL):



The screenshot above shows an override that will operate on all images that are located on the `/res/img` path and override their Cache Headers to the following default values:

```
cache override: public
max-age: 86400
cookie override: clear cookies
```

This way, the resources become publicly cache-able on various independent nodes on the network. Depending on the location of the caching node and the pathway of the request, the content will be served from caches instead of going through the proxy pipeline.

The result is a considerable lightening of server load, and a decrease in the number of page requests that go through the proxy.

Important Notes!

- The feature is meant to be used in specific cases with binary/static resources that are unlikely to be changed quickly and often. Although any type of resource can be marked as publicly cache-able, these settings should not be applied to dynamic content (be warned of the **likely discrepancies** between the original and cached content!). **Careful Consideration is Required!**
- Since Public Caching relies on the network to offload the burden of serving resources, changes take time to propagate. The `max-age` directive of the Cache Header indicates how long a cached instance of a resource will stay cached - the default is 24 hours. When the time declared for `max-age` is up, the resource is cached again via a request for the original.

Advanced settings

The Advanced settings screen is the *nitty-gritty* of the technical side of the proxy, the various features of which make complex project management and involved content extraction/management possible.

Some, such as **Freezes**, are general enough to warrant close attention from all users, others, such as the **Tweaks** checkbox list are required only in specific cases, so you can go through many, many projects without ever having to worry about them. In this section, we take a look at all options in the order that they appear on the Dashboard.

Pattern matching: there is often a great deal of text on a website that is not targeted for translation, because it is highly repetitive or numeric in nature, usernames, timestamps and prices fall into this category of content, to mention a few of many possible cases where Pattern Matching might come handy.

You can construct regular expressions and add them here, one regular expression per line.

If that regex has a capturing group, anything that is **matched by that capturing group** will be treated as **translation-invariable**.

It is important to note that if you add regexes here later into the project, your existing segments will remain in the Workbench. Pattern matching does **NOT** delete or exclude any segments.

Example: filtering usernames and dates

```
Posted by youTuTroll on Apr. 4, 13:22, 2010
Posted by joey0405 on Apr. 6, 03:29, 2010
Posted by cunning_linguist on Apr. 6, 03:45, 2010
Posted by bornToBe27 on Apr. 6, 10:01, 2010
Posted by OMEGA on Apr. 7, 11:59, 2010
```

Repetitive and inconsequential from a translation/word count viewpoint, the usernames and dates should be made translation-invariant in the content above. You can add a regex for it (such as the one below) in the textbox and Easyling will exclude whatever is matched within a capturing group.

```
Posted by ([\w]+) on ([\w]{3}\.\s[0-9]{1,2},\s[0-9]{2}:[0-9]{2},\s2[0-9]{3})
```

We recommend that you visit <http://www.regex101.com> and test your regular expressions with example snippets before setting them on a live project.

Freeze

The following are important options for when translation has already begun in earnest on a batch of content. At this point, it is a good idea to **freeze** the page list and the segments - that is, prevent *inadvertent* addition of new text or pages to the list you are currently working on.

WARNING: These settings can be overridden via explicit user action: adding Pages via the blue icon in the page list, or running Scans with the appropriate settings in place.

As you based quote on a specific Discovery/Scan (in other words, a specific state of the site at a given time), it wouldn't do to return a few days after winning the project only to find that the news section was expanded with two new items and you have 4500 new source words to deal with - indelibly added to the project, but not part of the original deal.

The following features are available:

Freeze pagelist: Prevent new pages from being added to the pagelist. If you turn this option on, you can use the various preview modes without having to worry about new content.

Handle unknown pages as externalized: Pages that are not already in the pagelist will be *externalized*, that is, not translated *at all*, the same as if the page was manually excluded. Be aware that on a live site, this may result in an SEO penalty (due to duplicate content being detected by the crawler)!

Freeze translation memory: No new translatable segments will be added to a project as long as this Freeze is turned on (automatically enables Page Freeze).

Group pages

Specify path rules to group pages into one entry, preventing new pages from being added, but not removing preexisting pages, and making the grouped pages share a single dictionary, necessitating translation of only one. The pages will be grouped according to the path rules specified in the textbox, one path per line.

This feature is not meant to and does not decrease the amount of pages a Discovery has to crawl, but it makes project maintenance easier, as it prevents your page lists from being overcrowded with repetitive URLs.

JavaScript translation

This field contains the capture group definitions used to extract attribute-value pairs from JavaScript files selected for translation/localization. After entering the capture parameters and re-crawling the site, the proxy will display the selected JavaScript files as translatable pages in the pagelist, from where they can be selected for translation in the List View like regular pages, and any values for the selected attributes will be made available as translatable entries, which are treated identical to regular entries.

Entering “html” (note that the switch is separated by a space!) after the path specification will result in the proxy applying its HTML parser to the match instead of a plaintext parser, stripping out HTML markup and only offering the actual content for translation (otherwise, should the match contain markup, the translator must take care not to alter it, or risk breaking the translated site).

If a field of the JSON being parsed contains further JSON data in a stringified form (“key”: “{\\\"key\\\":{\\\"key\\\":\\\"value value value\\\"}}”), the path can be passed to a recursive JSON translator by appending “json” to the path, then extending the path on the next line by adding “.json.”.

XPath Translation

The proxy can translate XML (eXtensible Markup Language) files sent by the remote server, according to the XPath standard of specifying elements of the XML structure. Similar to JavaScript translation, entering the “html” switch will result in the HTML parser being applied, while no switch will parse the match as plaintext.

Mark multiple Resources as Translatable

Using URL prefixes (N.B. fully qualified URL prefixes, including protocol, host, and possibly path structures!), the Proxy can enforce dictionaries over multiple resources in a single rule. This is especially useful if the site under translation contains an API (especially CREST APIs) whose responses also require translation, and each endpoint is served on a different path; in this case, entering the root of the API here will automatically capture all responses from that path without having to individually mark them as translatable from the Resources menu.

Additional Remote Request Headers

If the remote server requires certain headers to be present to serve legal responses, it will not be crawlable by default, as the crawler will not supply these. Entering the required headers here will result in them being appended to every request sent by the proxy, including the crawler requests, letting you crawl the site as required.

Ignore Classes

If a certain class of elements are not ought to be translated, they can be entered here. Elements with these classes will be treated as if they had the `<translate=no>` attribute, and will be treated as translation-invariant.

Ignore IDs

Similar to the “Ignored classes” option, this allows the treating of specific elements as translation-invariant, this time through HTML IDs.

Process Custom HTML Attributes

Some Content Management Systems may employ non-standard HTML attributes on various elements to style the page or otherwise affect aspects of their operation. If some of these attributes contain translatable text, you can enter them into the “As text” field to instruct the proxy to extract them. If they contain URLs that need to be mapped to the translated domain, you can use the “As link” field to instruct the proxy to map those non-standard link elements as well.

On request, it is also possible to activate an HTML parser for certain attributes, should they contain HTML formatting in their values.

Tweaks

In this section, you’ll find checkboxes for settings that apply to very specific circumstances. For those special snowflakes and occasions. When in doubt, contact us!

- **Retaining original DOCTYPEs:** By default, the proxy generates an HTML5 standards-compliant file to send to the client. If, for some reason, this causes problems due to the site relying on HTML4 standards for operation,

some of which may be deprecated by HTML5, enabling this option will cause the Proxy Application to retain the original DOCTYPE declaration of the source page.

- **Determine document type by GET instead of HEAD:** some servers may return different responses to the HEAD request we use to determine document type than the GET request used to download content. Enabling this option forces the proxy to use GET requests for all operations, getting the correct content type from the server (as far as server-side configurations will allow).
- **Detect JavaScript content type:** JavaScript may not be explicitly declared as such on the server, being sent to the client with misleading content types. This causes the proxy to mis-identify such streams and not offer operations reserved for JS files. Enabling this option will force a deep check on each file sent to the client, to determine if they are indeed JavaScript code, regardless of their declared content type.
- **Download images through the proxy:** this will instruct the Proxy Application to attempt to map all `` attributes to the proxied domain. This is especially useful if the images are served only after authentication.
- **Attempt to place tags according to punctuation when using TM-based pre-translation:** if using a TM-based pre-translation, the Proxy Application may encounter segments where it cannot replace the XLIFF tags automatically, due to overly large differences between the contexts (possibly because of a changed site). If this option is enabled, the translator will try to replace the XLIFF tags according to their positions relative to punctuation marks in the segment. If successful, the TM entry's confidence score will be increased by 0.1.
- **Translate excluded pages when viewing them in Preview mode (but still not in live serving mode):** It may be necessary sometimes to view excluded pages as if they were translated, in order to assess their layout, without actually making them available on the live site. Enabling this option allows just that, by propagating translations to the excluded pages if viewed in Preview mode, but keeping them untranslated in Live serving mode.
- **Translate javascript : attribute:** The Proxy is capable of extracting and translating code from the onclick attribute of elements. This feature may be used when a site uses this attribute to store translatable content inlined into the attribute and requires this content to be translated. Currently we only process the onclick event's contents, all other events in the javascript attribute will be ignored.
- **Detect and handle JSON-in-string responses, like " {\ "ResponseCode\ " : \ "BadRequest\ " } :** string-escaped JSON-format responses can be handled by activating this tweak. If active, the proxy will first attempt to string-unescape the response before passing it to the JSON parser, in order to recreate its base form.
- **Make content in `<script type="text/html"></script>` translatable as a whole (don't try to parse it as HTML):** Script blocks may contain template data requiring translation, which is often signified by the text/html content type (instead of the more usual application major type). In such cases, HTML parsing can be undesirable, and can be bypassed by activating this option.
- **Send a canonical link http header pointing to the original page on externalized pages:** the proxy can insert a Link header into externalized pages, in order to avoid the SEO penalty associated with duplicate content. This header will point to the original address, and have `rel=Canonical` added, to designate the relationship.

Manual Publishing

Manual Publishing is advanced project control feature that gives project owners the ability to withhold the translations from being published on the live page (but not the preview, as it will always display the latest translation available!) until further notice.

The feature can be activated from the Advanced Settings page. Once activated, it will affect all translations going forward, but already-existing ones will not be “un-published”.

Once active, a new item will appear in the Bulk Actions menu of the Workbench: “Publish”. Running this action will cause all selected segments to be synchronized with their displayed translations, and once the action finishes, the markers in the status bar on the right of the entries will change to reflect this. If the action encounters an error, the

server will attempt to rectify this by publishing the entire entry, after confirmation from the user. If the error is not recoverable, it will list the segments in error.

Default segment state

By default, the proxy will add new segments during a crawl as “Approved”, making them available for translation immediately. However, if the user/client so desires, this behavior can be changed to adding new segments in one of the two other states, “Pending” or “Excluded”. If the default setting is changed, the project owner, backup owners, or users with the Customer Role can alter segment states.

“**Pending**” segments are those that are awaiting a decision on translation. They will not be included in exports unless the relevant option is selected at export-time, and they will not appear in the Workbench for translation unless filtered for specifically. Those users able to alter segment states may change the state of these segments either by approving them for translation or excluding them altogether.

“**Excluded**” segments are those that have been deemed as not requiring translation at all. Unless the relevant option is selected, they are not included in exports and will not appear for translation unless filtered for specifically. Excluding a segment is not final, however: those users able to alter segment states may approve them for translation, making them available again.

Translation memory

Easyling’s translation memories (TM) are used to store existing translations for segments. You can import translations from *.tmx files or populate the translation memory with translations you have added in the Workbench.

Type in the name of your Translation memory and click on Create. Your TM will appear right above the input field, along with the “Add target language” option. You can also use the dropdown menu to assign TMs you’ve created previously.

Hovering over a translation memory reveals a menu, which allows you to import/export *.tmx files or populate the TM with existing translations. Use the “Remove” button to delete the TM.

The screen is furnished with a search field you can use to look up segments in a TM.

NOTE!

The TMs you manage on this screen and the Workbench are different from the Translation Memory that is billed: they are used for concordance lookups and automatic translation, and they are completely free to create and use.

On the other hand, the proxy TM associated with a project is a database in the cloud that holds translations of segments. When you reimport your XLIFF files or run auto-pretranslate, you are writing to this project TM, which is billed the first time around.

Pre-translate

By using the Pre-translate menu, you can initiate a process which supplies untranslated segments with a preliminary translation in order to mitigate bleedthrough on the proxied pages. You can use pre-translate with a Translation Memory or a Machine Translation API of your choice. Easyling supports Google Translate, iTranslate4eu, Bing Translate and GeoFluent.

In order to use Machine Translation, you need to go into your Account and set up your MT account Ids/authorization keys in the “Machine Translation” menu.

Auto pre-translate

Automatize pre-translation: you can set up the project to automatically execute a pre-translate process for any new content that is found - you can hook it up with your TMs, or any Machine Translation services you subscribed to.

Pre-translation of incoming content can be done without user intervention or oversight. If new content is encountered, and at least one source is configured, a user-configurable timer starts counting down. Content is collected while this timer is running, and once it reaches zero, translation begins from the configured sources.

Content that cannot be translated (if no matches of acceptable confidence level were found in the TM or if the MT-engine returned no translation) is packaged into a Work package that you can have sent as XLIFF files to external systems, such as XTM, XTRF, or your Dropbox account. There are links at the bottom of the dialog you can use to visit the settings screens where you can set up these accounts.

Remove project



This is always displayed in red. By clicking on it, you remove the project from the datastore: translations will no longer be served and the translated domains will return an error to browsers.

Be careful!

3.2.2 Discovery

General

So, what is a Discovery?

Before you can translate a website, you need to **extract** all content from it using the Scan function. A Scan, however, writes to the database, which, as you know from our pricing model, will cost you money. In that sense, it would be

risky to let loose an unlimited Scan on a website you're not familiar with. To forego Content Extraction and limit the crawler's actions to finding out about the structure and word count of a given web site, a more tentative approach is needed: a Discovery.

Discovery

Dashboard

Discovery

Pages

Resources

External links

Content

Page modifiers

Publish website

Wallet info

Owner:

Valid until:

Credits:

Current plan:

Billing period

Views: 0

Words added: 0

Words translated: 0

Generate report

Website discovery

Discovery collects the URL list and the word count for all individual pages of the website. Easyling will visit all the pages and follow all the links within the original domain in order to show you where the content is. As a next step, you can **exclude individual pages or entire folders** (by their URL path prefix) to get a final word and repetition count. Finally, proceed to **Content** to download the text for the selected pages only.

Settings

☐ Process pages in source language only (requires Google language detection API)

Ignore queries

Group pages by ignoring query parameters

Matching source segments are merged across pages in the group

Discovery cookies

These values are not saved for security reasons. [More about cookies](#)

Discovery page limit number of pages maximum Leave blank for Unlimited

[view history](#)

Discover now

The process maps the structure of the site by scanning each page for link elements and trying to follow these links. The content of the website is not stored, only the URL address of the visited pages, together with their status info. Any page that is verified to exist is marked as Discovered, and the ones that returned an error message (most commonly redirection (HTTP301-302) and page not found (HTTP404)) are marked Unvisited in the page list. For more information on HTTP status codes, see [here](#). Although this process doesn't extract content, it provides a preliminary wordcount and a repetition rate as well. It has a cost of 1EUR per thousand pages.

IMPORTANT!

If you exclude pages during discovery, changing the rules only will not include them in the new discovery; you have to add them manually through the “Add pages” dialog.

Your primary choice is between limited and unlimited Discoveries.

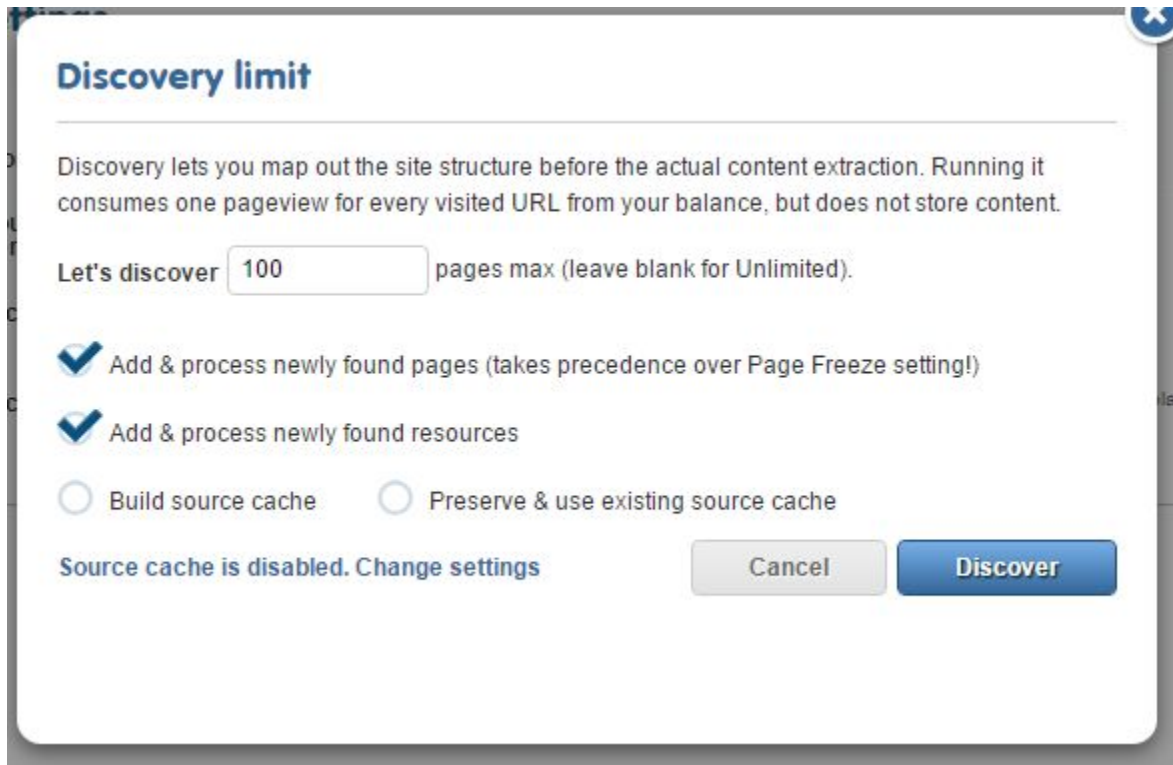
A limited Discovery is the best way to “get to know” a website - if you set a reasonable safety number, such as the default 100, you can be sure the process will go overboard when it finds an enormous forum or a structurally complicated web-shop.

Based on what a limited Discovery tells you about a site, you can set **prefix exclusion rules** in the page list or **cherry-pick** unnecessary pages for exclusion.

Wordpress sites, for example, tend to create links such as “/?page=2332” for every page, and it is obvious from the get-go that there is no real content behind these links. A limited Discovery will inform you of their presence, so that you can exclude them at.

By increasing the page limit and running Discoveries in succession while adjusting your inclusion/exclusion rules, you can work your way through a site structure incrementally.

Of course, if you have a clear idea of the size of the site, you can always set up an unlimited Discovery, wait for it to finish and set up your exclusion rules afterwards.



Discovery limit

Discovery lets you map out the site structure before the actual content extraction. Running it consumes one pageview for every visited URL from your balance, but does not store content.

Let's discover pages max (leave blank for Unlimited).

☒ Add & process newly found pages (takes precedence over Page Freeze setting!)

☒ Add & process newly found resources

☐ Build source cache ☐ Preserve & use existing source cache

Source cache is disabled. [Change settings](#)

Once the discovery is ready, you will receive an e-mail notification and the statistics will show up on the Discovery page. Based on this you can give a rough estimation for the website translation cost - both in time and money.

Statistics

	Segments	Source words	Latin characters	Latin characters with spaces	Asian glyphs
Total	2,106	10,290	54,520	63,104	0
102% repetitions	1,642	6,091	32,165	36,891	0
101% repetitions	15	186	962	1,136	0
100% repetitions	18	194	978	1,156	0
99% repetitions	18	98	492	575	0
98% repetitions	0	0	0	0	0
Unique	413	3,721	19,923	23,346	0

Referred domains
Number of referred domains: 11
Show list

JavaScript generated content
Check

To understand the way statistics work in Easyling, we have to start with the most important fact: in our application, the largest unit of measurement is a block, which is usually represented by a `<p>` or a `<div>`. Blocks break down to segments, segments to words, words to letters. Since Easyling deals exclusively with content in webpages, HTML tags also play an important part in weighing the repetitions.

It is also important to note that the statistics in Easyling are different degrees of repetitions. The website's content is repetitioned against itself, simulating a translation process, not unlike the Homogeneity feature in MemoQ.

With that in mind, here is a breakdown of the percentages in our Statistics. You can find these explanations if you hover your mouse over each repetition row:

102% - Strong contextual repetitions: These are block repetitions. Every segment in the block is a 101% repetition, and all the tags are identical. We do not charge for these repetitions and they are propagated automatically within the project.

101% - Contextual repetitions: These repetitions are comparable to the 101% repetitions in MemoQ, or Context Matches in SDL Trados Studio. Both tags in the segment, and contexts (segments immediately before and after) repeat.

100% - Regular repetitions: This one is straightforward, and comparable to the Repetitions count in MemoQ or Trados. The segment is repeated exactly, including all tags.

99% - Strong fuzzy repetitions: In this case, a repetition is found after few transformations on the segment before comparing: tags from the ends are stripped out, words lowercased, numbers ignored.

98% - Weak fuzzy repetitions: Here, all tags are stripped out, not just the ones in the end; words lowercased, numbers ignored.

You will find an illustration of the various repetitions in the table below. Hover your mouse over any of the matches to highlight the differences.

Discovery settings

Fine-tuning Discoveries and dispatching them on a site is the bread-and-butter of Easyling, and arguably the most important part of the setup-phase of a project. In this section, we take a look at the various settings, convenience features that are at your disposal. Some of the most important overarching settings will be displayed in the main screen of the Discovery screen, others will be revealed by the Discovery dialog, which opens after you click on the blue **Discover now** button. Let's take a look at each of those.

Discovery Screen

Ignore queries: Add any query parameters to this field, separated by commas, to ignore them during Discoveries.

By default, the following URLs will turn up in the Page List as two separate pages:

```
http://www.example.com/examplepage.html
```

```
http://www.example.com/examplepage.html?querystring_1=foo&querystring_2=bar
```

Query strings can influence the content that will be served, but it is also common for a page to be completely unchanged regardless of their presence or absence. In such cases, query parameters are somewhat bothersome: they cause the Discovery to revisit the same page multiple times, they overcrowd the page list with unnecessary entries and divert attention from the real page that holds the content to be translated. It is a good idea to trim the list and exclude these superfluous pages using this feature.

Group pages by ignoring query parameters: In the same vein, you can also group pages based on ignored query parameters by adding those query parameters to this field. Query parameters are bona fide part of the URL, and therefore may or may not induce changes on a case-by-case basis. For every changed query parameter, we create a new page record. If you are sure about a set of query parameters not inducing any changes in page content, you can use this function to prevent them from getting picked up.

Discovery cookies You may extract the cookies from a user session and pass them to a Discovery. This feature is most useful when you need to deal with login walls and content that is only revealed within the context of a user session.

Note that we cannot guarantee that the cookie extraction method works with all proprietary logins, as there are several solutions which can use additional security measures such as IP checking, or checking the User-Agent header. As all the requests from the proxy and the crawler come from the Google Cloud, it is necessary for the target site to not block them. The success of a Discovery via the use of session cookies is highly dependent on the implementation details of a site.

Discovery page limit: It is usually not a very good idea to unleash unlimited Discoveries right off the bat, not unless you have a 100% understanding of a site's structure. Granted, Discoveries don't really cost much, so you are in no danger of incurring any serious expenses. But a Discovery can take time - and the shorter the time span required to create a quote with high confidence levels, the better.

For this and other reasons, it is a good idea to start with a "feeler" Discovery and take it from there, adjusting query string settings, adding prefix-based exclusion rules for unnecessary subsections, fine-tuning the settings as needed - this way, you can safely shape the Discovery incrementally, until it visits only those pages that you deem necessary.

Discovery history: By running subsequent limited Discoveries using progressively refined settings, you will also get contextual clues about site state by checking the Discovery History. By clicking on the small **view history** link under the timestamp of the last Discovery, you can avail yourself of all the statistical information revealed by your previous Discoveries by using the "Show info" button in the opening dialog. You can also verify the status and end state (termination reason) of your Discovery processes here.

Referred domains Links pointing to a different domain than the project website address will be treated as external referred domains and listed here for your convenience. Note that specific instances of external links (such as links to streamed media) can be localized(replaced) over the proxy. See Content > External links for further details.

Discovery Dialog

Discovery limit Repeated from the main screen, you can limit the Discovery to a set amount of pages before exiting.

Add & process newly found pages any pages the Discovery process finds during a crawl will be added to the page list. Note that this is explicit action by the user, therefore the Page Freeze setting has no effect in this case.

Add & process newly found resources Instruct the Discovery to add newly found Resources, such as *.js files to the project. The proxy makes collecting Resources a trifle, and there is a plethora of in-depth methods to translate content embedded within them. But translation of Resources is also known to be one of the trickiest parts of website localization. Beware, and woe betide the unready!

Add & process newly found image resources By enabling this option, image content found on the site is added to the Resources screen, where you can add localized counterparts and instruct the proxy to serve those in place of the original.

There is often a great deal of image content on websites, so indexing all of them can take a very long time to finish; it is a good idea to turn this option off the first time around, and enable it only later, when it is clear that image localization is also going to play a part.

Skip content type check if already determined (to be added)

Limit crawl depth There is more than one way of limiting the scope of a Discovery. In scenarios where you have to Discover richly interlinked pages (a wiki, for example) you can use the crawl depth to limit how deep the Discovery should be allowed to follow links.

Ignore pages matching the following regular expression You can specify any regex to apply to all URLs and ignore the given page if there is a match. This is a user- and crawl-specific setting. As always, we recommend that you test your regular expressions on regex101.com before using them. In this case, you can export your page list from Discovery > Pages and use that list as a test string for your regex.

Limit number of simultaneous requests Control the number of requests a Discovery is allowed to send to the original site simultaneously. Use this function to prevent the Discovery from flooding a smaller server with requests.

Exclusion rules You can set Discovery-specific exclusion rules by selecting *Use temporary exclusion rules* from the menu and adding your prefixes to the field that opens - use the buttons on the right to choose whether the prefix entered is an inclusion or exclusion rule.

By default (*"Use current exclusion rules"*), a Discovery will use the exclusion rules that you've set up on the Pages list screen.

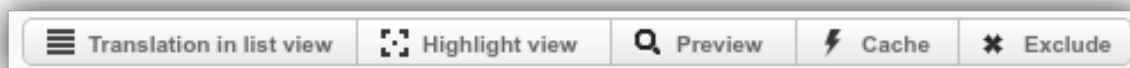
Build Source Cache Source Caches have to be built during Discovery or Content Extraction crawls. This checkbox becomes available after you have enabled Source Caches in Dashboard > Page Cache. A dropdown menu will also appear, where you can select which of the Source Caches you'd like to use for the current Discovery. Please see the Source Caches section of this manual for the details. The option to keep the current Source Cache intact (*"Preserve & use existing source cache"*) is also available - these two options override each other, you should turn the latter off when you're building a Source Cache.

After clicking on the blue "Discover" button, the Discovery starts and you will be returned to the main Discovery screen, where you can follow the word count statistics and check the status of the currently running process.

Pages

The Page List is accessible from both the Discovery and Content menus. By default (before starting any Discoveries or Scans), only the root path ("/") will be present. Later, all pages/paths added manually or found during Discoveries and Scans will be listed here.

Clicking on a pathname will take you to that URL on the original site. If you hover your cursor over it, a menu will appear to the right:



The first two buttons, “Translation in List View” and “Translation in Highlight View” will take you to the workbench. To learn more about the various editing modes, please visit the **Workbench** part of the documentation.

NOTE! Most of these buttons are disabled and unclickable as long as no target language is selected in the Language dropdown on the left side of the Page list!

By clicking on the third button, “Preview”, you will be introduced to the preview proxy mode, which you can use to review your translation efforts in the full context of the original website over the proxy. By holding Ctrl and clicking on the Preview button, you will be taken to another useful mode, the x-proxy. We’ll get into these proxy modes in detail later on in this manual (see **here** for the details).

At the top, you will see a few icons you can click to access detailed function accessible from the page list. They are as follows:

Copy current URL list to clipboard will open a new dialog containing project information and all pages in a pre-selected list. You can copy & paste this list, or click on Export to receive an XLS spreadsheet in an e-mail.

Search bar When the page list gets too long to scroll through manually, you can use the Search Bar to look up any path names that match your query. This is a forward-search (note that regular expressions are not supported in the Page list).

Resources

Resources are binary content, such as images, PDFs, CSS and textual sources such JS files, etc. It is important to remember that the content of these resources is not extracted for translation, so you have to translate / edit them separately.

When you open this screen, a subset of all Resources collected during a previous crawl will be listed. Scroll down to load more, or use the search field for lookup (as some websites are quite Resource-heavy, the search function is recommended).

Hover and click over any Resource to display a localization dialog for it. You can click on the green “plus” icon to mark the Resource as translatable. After doing so, a few fields will be added to the dialog: a source and a target language text box you can use to transcribe the Resource, and a target URL field which will contain the the URL of the localized Resource on the proxy. Use the “Choose File” button to upload a localized Resource.

Note that you can only localize Resources *one at a time* and on a *strictly target language basis* (make sure you have the appropriate target language selected on the left side of the screen).

You can click on the flags on the lower left side of the frame to check the localized and original versions of a Resource (especially images, which will be displayed within the dialog). Document browsing, however, is not supported on the Resources screen.

External links

All external links will be collected (and all external domains will be listed in Discovery), and sometimes, external links need to be localized, such as when documents, data files and downloads are served from an external domain. If you are translating `www.example.com`, but it’s downloads section is filled with links to files on `files.example.com`, you can use this feature to replace the links to point to a localized Resource of your own choosing.

3.2.3 Content

General

The main difference between Discoveries and Scans is that Discoveries do not extract or store content. They are used for statistical purposes, such as providing a word count or mapping a site's URL structure.

A Content Scan, however, writes into the database. It extracts and stores the source content to be translated. New source words added to the database are billed, therefore care must be taken when starting Scans. If you are not really sure about a site, stick to Discovery until you gain better understanding of its structure! Unlimited Scans especially require attention: they will relentlessly add all content to the database according to the specification you set for them.

Settings you used to experiment with Discoveries can also be used to initiate Scans - for a detailed explanation of the different settings, please take a look at the Discovery chapter of this section.

There is button on the main screen that you will not find in the Discovery menu, that is "Update word count", which you can use to refresh the statistics window during or after an extended Scan.

Pages & Resources & External links

These screens are duplicated between the Discovery and Content menus for ease of access. See the pertaining documentation of them in the Discovery part of this section.

There is one major difference between the two lists: the exact behavior of the Add Pages button depends on where it is clicked, since it will start the appropriate crawl. If the dialog is opened on the Discovery-pages screen, a Discovery will result; summoning it from Content-pages will start a Content Extraction crawl, generating potentially greater costs!

Manage segments

By clicking here, you will be taken to the Workbench on a new tab. The Workbench is the working end of the proxy service - please visit the Workbench section of this manual for the details of how to use it.

Work packages

Work package generation is the recommended method of dealing with exports/new content. Along with the general import and export features, work package generation becomes available after the first round of Scanning (Content Extraction) has finished and translatable content (segments) has been placed in the database.

Initially, the screen will be empty, but if you click on the blue "plus" icon, a dialog will pop up where you can generate new work packages using a variety of settings at your disposal. Let's take a look at them one by one.

Work package generation

Work package name You can name your work package any way you like. As a suggestion, it is recommended that you give it a name that accurately reflects its contents, so as to make project management using work packages easier. So, for example, if you generate a work package for Arabic as a target language on April 10th, 2016 for all new content, you would name it something like *ar-SA-untranslatedContent-2016-04-10* and so on.

Source language The source language for the project is displayed in this field (settable from the Dashboard main screen).

Work package language You can add multiple target languages for the Generation Process, but remember that a separate Work Package will be created for each target language.

Timeline You can specify which Content Extraction cycle you'd like to see in the Work Package that is to be generated. Click on "change" and use the timeline window to specify the time period, and the generator will limit its attention to just that timeframe. Select any two timestamped buttons to define the start and end of the timeframe.

Filter pages You can use your current selection (set on the Pages list screen), or define specific selected pages (both paths and full URLs are allowed) or folders using path prefixes. When you use a predefined list of pages (selected pages) the additional option of including both the HTTP and HTTPS versions of pages will be displayed as a checkbox.

Options Miscellaneous options you can use to fine-tune the properties of your Work Package.

Split your package at preset words to create easily manageable chunks to assign to translators. Disable it to put everything in one big file (Note, however, that Work packages will be automatically split at 30.000 words)

You can choose to create separate work packages for hidden elements, and set the Work package generation to automatically take care of exporting for you. As with the general import function, you may elect to export only those entries that have no translation yet.

If you configured XTM, XTRF or Dropbox to integrate with the proxy, you may instruct the software to automatically send the exported XLIFF files to either one of those services. Use the dropdown menu to make your choice (only available if the checkbox is checked).

Work package list

You will receive e-mail notification about your generated work packages and they will appear in the list. You can use the refresh button to get an up-to-date view of your work packages.

If you hover over a work package, a floating menu will appear to the right with two options: XLIFF export and Progress report generation.

Import translation

When your XLIFF files have made a round in your favourite CAT tool, it is time to reimport them, which is a straightforward process - click on "Import translation" and the Import dialog will pop up, where you can click on "Choose file" to browse for your translated XLIFF files. Click on "Import translation" and wait for the import process to finish.

You may optionally set the workflow status of imported segments to Translated-CONFIRMED, and you can use the second tab in the dialog window to reimport translations from your Dropbox account.

NOTE! Importing is not instantaneous, it may take some time to finish depending on the size of your XLIFF!

Export translation

By clicking on this link, the general Export dialog will open. Most of the options will be familiar from the Work package generation dialog. Choose a file format of your preference (with XLIFF being the *de facto* industry standard, it is recommended that you go with this option, as it is more structured than CSV).

Languages to export You can choose to export more than one target languages in one go, or choose to export them all.

Export range Click on "change" and use the timeline window to specify the time period, and the export function will be limited to that timeframe. Select any two timestamped buttons to define the start and end of the time frame.

Export You can choose between exporting only those entries that lack a translation, or include all entries.

Unique segments only Check this option to ensure that identical segments are not duplicated in your exported XLIFF file.

Skip excluded and pending segments from export Export only those segments that have been approved for translation on the Workbench.

Trim export to contain as few tags and whitespaces as possible Use this option to clean the source segments of tags for the purposes of translation.

Split enter the number of segments that you'd like to have in a single file, and the export will split your segments into separate files accordingly.

Copy source to target where empty Some CAT-tools require a target entry on import. Activating this option will copy the source content to the target entry during the export, so the tool can be used to edit it.

Send XLIFF files Choose among your XTM, XRTF or Dropbox accounts to send your exported files to. Each of these accounts has a link at the bottom of the dialog that will take you to the given account's setting screen.

Click on Export to execute. Wait for the process to finish - you will receive an e-mail when your files are ready.

Previous exports

Although the notification e-mail does contain the link to your exported file, there is no reason to track it from your mailbox. This dialog will list and organize all previous exports for your viewing with the appropriate timestamp, target language and word count displayed right next to an Export list entry.

Currently running processes are also displayed, with a spinner indicating that the Export is still underway.

3.2.4 Page Modifiers

General

While translation over the proxy makes website localization easy, it is also a fact that UI issues can sometimes surface when a page is proxied.

Suddenly served in a different language from the one the UI designers anticipated, websites can display a variety of UI quirks. As website translation is still, sadly, more often than not an afterthought, these issues have to be dealt with in the translation pipeline.

Word length differences between the source and target languages might cause the menu to crowd. Differences between the lengths of different text blocks can cause otherwise well-designed CSS rules to behave in unexpected ways. Not to mention the plethora of plugins that run on a modern website - there is a big pool of possible glitches right there.

Fortunately, changing the site content over the proxy is relatively easy. The Page Modifiers function is made available for this reason: to empower you to add your own CSS rules and JavaScript snippets to influence the way a given proxied site looks to the user.

Because the datastream must pass through the proxy to have the translation embedded, the proxy can insert JavaScript modifiers, modify style sheets, even embed entire pages that do not exist on the original site.

CSS rules and JavaScript are injected into each page that is served over the proxy.

CSS editor

The proxy can be used to insert locale-specific CSS rules into the site being served. The rules are inserted as the last element of the head on every page served through the proxy. A very common use case for this feature is RTL conversion of a website: almost always necessary when one of the target languages is Arabic.

There are a couple of caveats:

It is good practice to make each CSS rule language specific (Easyling will always insert the appropriate locale code into the HTML):

```
html[lang="de-DE"] ul.one-selector li a {
    float: left;
}

html[lang="fr-FR"] .another-selector h2 {
    height: auto;
}
```

If you omit the `html[lang="fr-FR"]`, the CSS rule will be applied in all target languages, which might not be the behavior you expect.

JavaScript editor

Here is a partial, but plausible example of a JavaScript Override:

```
(function() {

    "use strict";

    document.addEventListener("readystatechange", function() {
        if (document.readyState === "interactive") {
            resizeTextBoxOnContactPage();
        }
    }, false);

    window.addEventListener("resize", function() {
        resizeTextBoxOnContactPage();
    }, false);

    // make function page-specific. Alternatively, add check to eventListener above
    function resizeTextBoxOnContactPage() {
        if (document.location.pathname !== "/contact-us") { return };
        // implementation details here
    }

    // more function definitions here

})();
```

IIFEs (Immediately Invoked Functional Expressions), such as in the script above, are a generally good design pattern for Page Modifiers, since the programmer can work within the confines of their local scope. The function and variable definitions will never escape into the global namespace - which is a conversely bad idea, as it could potentially conflict with original code on a site.

An in-depth discussion of JavaScript is not possible in this manual, but there is generally no limitation on the type of code that can be injected via Page Modifiers. **Security implications should always be considered when adding custom code to sites.**

You can also use jQuery if it is present on the original site.

Content override

The proxy can create a “virtual” page in the site or override an existing one with custom code. For any requests to an overridden page, the corresponding remote server request is not sent, and the override contents are used as the basis of

the translation. The source is not required to be HTML, custom content-types can be entered, along with customized cache headers, and status codes (HTTP status codes are restricted to those permitted by the Java Servlet class!) - note that the 300-family of status codes requires the Location header to be defined as well.

3.2.5 Publish Website

General

In this menu, you are provided with the means of publishing your website on a pretty domain after translation tasks have finished.

The publishing interface has two options to make the site available to the web at large: serving domain mode and subdirectory publishing mode.

Domain settings

This screen opens by default after clicking on “Publish website”. The following settings and fields are available:

Target language Choose which target language you would like to configure a serving domain for.

Temporary domain This is the preview proxy you can use to preview your translations live (this is immutable).

Access control you can add basic HTTP authentication to the proxied site in order to restrict access to it. Click on *Configure access control....* Select the proxy modes that you’d like to limit access to, enter your username and password. You can add multiple users to the list.

Serving domain mode (de.example.com)

In the serving domain mode, Easyling will publish the translated site either on a subdomain of the original (the default behavior, such as `de.example.com`), or on a completely separate naked domain (such as `example.com`). In order to use this mode, you (or the client) will have to modify the DNS settings corresponding to the original domain - the three to five records (three for subdomains, five for new naked domains) that need to be inserted in your DNS settings are found below the two input fields, and will change as you enter the desired serving domain.

CNAME settings There are three CNAME settings that are required on your domain to enable publishing of your website. Each of the lines in the table that is displayed has a specific function:

CNAME 1 allows mapping of the subdomain in Google AppEngine, enabling us to alter the datastream and translate the site. This row is computed from the serving domain entered in the second input field, and needs to be added once per serving domain.

CNAME 2 The second line determines where the currently selected target language will be published. This defaults to the language code, but you are not obligated to keep it that way. This has to be entered separately for every target language you publish.

CNAME 3 The third line verifies ownership of the original domain. This is computed from the user’s ID who is currently looking at the publishing page (i.e. different users will see different values, so one person should communicate this to the client and hit the Verify button), and needs to be added only once per serving domain as well. You can set the subdomain and domain where the translated site will appear.

After all the settings have been entered into the DNS records, there is short time while the changes propagate and are replicated across the world. This can vary wildly with the DNS and hosting providers, taking anywhere between one and twenty-four hours. It is recommended to wait out the twenty-four hours, as Easyling will not be able to save the published domain until all checks are passed.

HTTPS & SSL certificates Additionally, if the original site was HTTPS, or the translated sites will be served over a secure channel, an appropriate certificate and its associated key will also be needed. Ideally a wildcard certificate (one certificate for all subdomain), but Extended Validation certificates can also be used, although they require more setup work. Our support team can also provide you with the Certificate Signing Requests necessary for the generation of these certificates - this will also have the added benefit of not having to send the private keys over the wider web for recording in AppEngine.

See the pertaining section of the documentation here.

Subdirectory publishing (example.com/de)

The alternative to subdomain-based publishing is to publish the site to retain your own domain, and publish the site as a subdirectory of the existing domain, i.e. the translated pages will be available on separate paths under your own domain. However, due to the way the proxy operates, this will require a specifically configured nginx reverse proxy server to be prepared and placed in front of your own webserver.

An example configuration snippet is available on the publishing interface, but this should be treated as only a guideline, to be customized according to your specific setup!

There are a variety of options at your disposal when implementing subdirectory publishing on your domain. You may choose to use `nginx`, `CloudFront` or any similar solution to use as a reverse proxy.

Language Selector

As translation of a website is nearing completion, it is usually necessary to prepare a language selector to make the various languages more accessible. There are two options:

1. develop a custom selector in-house
2. use one of the proxy default selectors.

In all cases, the addition of a language selector will need to be done on the original server and then propagated to the translated domains through the proxy.

Proxy Solutions

To spare you the expense and time of developing a custom language selector, we have prepared two default solutions you can use. Each of these require the minimum possible intervention on the original site for deployment. Note, however, that seamless visual integration of the selector requires some fiddling with the CSS rules. The following two types of language selectors are available:

Sidebar The sidebar is an easy-to-deploy-and-use language selector (with limited customization options) that doesn't require any changes on the original site beyond the addition of a single `<script>` tag to the source code. The language selector is a hovering sidebar which displays the flags of the selected target languages.

The sidebar is an easy-to-deploy language selector that displays flags for each specified target locale on the left or right side of the website. The horizontal orientation and display names are settable.

Set-up steps

On the Dashboard:

1. Click on the Publish Website menu
2. Select "Sidebar"
3. Select "Left Side" or "Right Side" according to your preference

4. Specify the Target languages by enabling the last checkbox on the screen

Note: You may also declare custom language names to be displayed by using the dropdown selector or choose “Language” to display only the language, but not the Country.

On the original site:

1. Finally: copy the embed code from the textbox and add it to the end of `<head>` tag on the original site. With the tag added to the original site, a reload should reveal the language selector.

Dropdown The Dropdown is the more customizable option. One bit of additional adjustment in the original source is necessary, however: besides adding the `<script>` tag, a dedicated `<div>` tag is also necessary – since the dropdown is more closely integrated with the site structure (that is, it has to be positioned in relation to other elements), this `<div>` tag will be used to locate the designated location of the dropdown.

Setup steps:

On the Dashboard:

1. Click on Publish Website menu
2. Select “Dropdown” (default)
3. Declare a unique class name for the `<div>` that will contain the language selector. The default is `el-dropdown` but you may specify any class name that does not clash with any other classnames in the original HTML. For the purposes of this tutorial, the classname will be `el-dropdown`
4. Select a display template. You may choose between “Language (Country)”, “Language” or entirely custom names according to your preference.
5. Specify the Target languages by enabling the last checkbox on the screen.
6. Copy the embed code from the textbox.

On the original site:

1. Locate the position in the markup that you wish to insert the Dropdown
2. Extend the HTML markup with a `<div>` tag. Furnish it with the previously specified class. In this case, it would look something like this: `<div class="el-dropdown"></div>`
3. In addition, Add the embed code copied from the Dashboard to the `<head>` tag of your page.

Post-setup:

At this point, the language selector is displayed after reloading the page in your browser (the JavaScript referenced in the `<script>` should take care of downloading and rendering it). The language selector contains a menu button, a container and target language entries you can customize via CSS rules. The language selector is self-contained for convenience, so it uses in-line styling. Postfix your CSS rules with “!important” to make sure that they are applied.

RYO or Existing Selector If a language selector is already present on the site, by far the easiest method to integrate the translated subdomains/subdirectories is to add them to the existing selector (although it does require the development resources of the site owner). An important feature of the proxy is the “`__ptNoRemap`” class, which can be used to prevent the links within the language selector from being changed over the proxy. Adding it to the link `<a>` tags or other functionally equivalent markup will allow the proxy to keep these links intact. It is especially prudent to protect the original site’s link (the proxy should not remap the link to the original site in the language selector). You may also wish to use the `translate="no"` attribute in places where you want the proxy to ignore the textual content of the HTML element.

Wordpress Plugin

Clicking on this menu will take you to the Wordpress site where you can download the Easyling Wordpress plugin.

The proxy provides an option to install the Translation Proxy directly to the Wordpress CMS. In this case no Translation Proxy fee will be charged, but the foreign traffic shall be handled by the content owner's infrastructure. For more information see our blog or the WordPress.org.

Client-Side Translation

The Proxy also provides an additional translation engine in JavaScript. A project can be exported into a ready-made JavaScript file that couples with the translation stub provided to enable translation within the client's browser.

This mode of operation is not without drawbacks, however! Currently, binary resources such as images will not be replaced, nor will additional JavaScript/CSS modifiers be injected, as the site now bypasses the Proxy entirely. Additionally, Automatic Change Detection will be restricted to the recurring crawls set up on the project, for the same reason.

The Client-Side Translation offering was premiered on 1 June 2016, and is currently in open beta phase. If you would like to participate as an Alpha Customer, please contact our support team from within the Proxy!

3.2.6 Integrations

General

(to be written)

3.3 Cookbook

3.3.1 Whitelabel Setup

Note: This article will use variable values, that will change for everyone. These variables are written here in the UNIX style of `${VARIABLE_NAME}`. When providing us information, simply replace the entire construct (not just the name itself) with the data in mind. There is also “global” variable to keep in mind: `${APP_DOMAIN}` refers to the domain chosen in point #3 to serve Easyling on.

1. A €200 EUR topup / month (recurring): Easyling under your own brand name is special service, offered to customers who cater not only to one or two clients, but put their weight behind the punch and open up whole new markets with our proxy solution.
2. A one-time setup fee of €200.
3. A custom domain name: you will need a place to serve Easyling (as well as any previews) on. Generally, our clients settle on `app.${yourdomain}.com`, but we can use practically anything that comes to your mind - the only limitation is that we are unable to serve the proxy on a naked domain (for instance, `yourproxy.com`). Just keep in mind that once you settle on something, and we set up your branded Easyling, it becomes fixed, so your decision is final.
4. Two logos: one goes on the Dashboard, the other goes on the Workbench. They should be transparent PNGs, ideally, but we can use other file formats as well. However, their dimensions are fixed: the Dashboard logo is set at 200x62px, while the Workbench logo needs to be 109x44px.

5. An SSL certificate: Easyling uses encrypted channels to communicate on, and for that, we require an SSL certificate to be made out for the domain name of your choice, and any subdomains it may have - Easyling uses your “app domain” to serve previews until they’re published, so your certificate must be a so-called “wildcard certificate” (a type of SSL certificate valid not only for `app.yourdomain.com`, but also `*.app.yourdomain.com`). Certificate issuers are likely to request a Certificate Signing Request (CSR) for the certificate, which we will have to provide.

In order to generate a CSR for you, you’ll need to provide a few pieces of data related to your company, which need to be incorporated into the certificate. Please provide the following by replacing the fields (this should look fairly familiar to your IT department) with the appropriate data.

```
1. `countryName_default = ${COUNTRY}`
2. `localityName_default = ${CITY}`
3. `streetAddress_default = ${ADDRESS}`
4. `postalCode_default = ${ZIP}`
5. `0.organizationName_default = ${COMPANY_NAME}`
6. `organizationalUnitName_default = ${ORG_UNIT}`
```

1. The final step is to configure your DNS servers; and if you use Google Apps for `yourdomain.com`, the setup process will require someone with Google Apps admin rights as well. You will need to add the following CNAME records to enable Easyling on your domain:

- (a) `${APPENGINE_KEY}.${APP_DOMAIN} CNAME ${APPENGINE_HASH}` - these values will be provided to you.
- (b) `${APP_DOMAIN} CNAME ghs.domainverify.net`
- (c) `*.${APP_DOMAIN} CNAME ghs.domainverify.net`

2. In order for Easyling to be able to send emails from under your domain, you will need to provide authorization to the email service. This is done by adding specialized DNS records.

- (a) `mandrill._domainkey.${APP_DOMAIN} TXT v=DKIM1; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQCRLHiExVd55zd/IQ/J/mRwSRMAocV/hMB3jXwaHH36C`
- (b) `${SELECTOR}._domainkey.${APP_DOMAIN} TXT ${DKIM_KEY}` - these values will be provided to you as they are domain-dependent.
- (c) `v=spf1 include:spf.mandrillapp.com include:sparkpostmail.com ?all;`
or if you’re already using an SPF record, add `include:spf.mandrillapp.com`
`include:sparkpostmail.com` just before the last operator.

3. Finally, if you want, you can specify the following information to customize the white label experience (this is completely voluntary and can be changed at any time):

- (a) name: Name of your branded product
- (b) greeter: Person signing the greeting emails for new users
- (c) team: Team name
- (d) greeter address: Email address of the person sending the greeting emails
- (e) greeter display name: Name of the person sending the greeting emails
- (f) noreply address: Email address used for automated emails
- (g) noreply name: Display name used for automated emails

(h) quote and wordcount signer: Person signing the quote and word count emails

Once all seven main points are settled, and we have the logos and the SSL certificate with us, we'll set up the white label version for you, and you'll be ready to start cracking your target market wide open.

3.3.2 Resource Translation

Resources are binary content found on the sites, suchg as images, PDFs, CSS and JS files, etc. Please note that the content of these resources is not extracted for translation, so you have to translate / edit them separately.

Replacing Images with Localized Versions

You have the option to replace images, downloadable files and other resources with their localized version.

1. Please make sure that you see the appropriate target language selected on the left side menu.
2. Navigate to the original image in the Resources view of the Content (or the Discovery) section.
3. Hover over the thumbnail image. A green '+' icon should appear.
4. Click this icon. It marks the image as 'localizable', i.e. a candidate for replacement.
5. Select the replacement image and upload it. It will immediately replace the original image, and the new image will show up on te translated site.
6. Check the Preview to see if the image was replaced properly.
7. Repeat these steps for all images and all target languages that need localized versions.

Please note that you can replace only 1 resource, for 1 target language at one time.

Extraction Issues with data-image Attribute

Images might be inside a "data-image" attribute in the source code. In such cases you have to set the "data-image" attribute up to be translated as a link in the Advanced Settings screen. The `` attributes are recognized by default as URLs, while data-* attributes are not, as they can contain any sort of data, and must be configured manually to tailor the behavior to the current project's needs.

The image might also be drawn from a subdomain. In this case you have to create a second project for the subdomain, mutually link the two projects together, and add the resource in the subdomain's project. In this case you have to publish the two projects together, in order to preserve the mapping.

3.3.3 Left-Right conversion

It depends on the site...

1. Best case scenario:

```
html {  
    direction:rtl;  
}
```

```
<html dir="rtl">
```

If it looks mostly OK, then all is left just some minor CSS fixes:

- flipping images (in carousel / slider)

- list elements' bullet is defined using a background image
 - If text's align is defined explicitly (e.g. using WP's text editor) inline, then each and every element must be overridden using `!important`
1. Bootstrap and framework alike:
 - some framework might provide a dedicated RTL css file [bootstrap-rtl.css](#)
 1. If this is not case, each and every element must be positioned individually

Mixed content within text

- When it comes to actually render the text (numbers), the direction is determined by a couple of rules. Please read [this](#) for the details.
- ***As a rule of thumb: during the translation (from LTR language to RTL), don't use your CAT tool to change the order of the numbers and text where the translation is with latin characters. Phone numbers like 1-800-123-1234 should be left in this order.***
- To make sure numbers are rendered properly at the end, a Left-to-Right (LRM) must be inserted before every number. The dash between numbers split them, so LRM must be inserted after them again. [Click to read more on how to insert these LRM characters.](#)
- The same holds for parenthesis, etc. Make sure you understand the rules; sometime LRM must be inserted *after* the closing parenthesis

```
<p style="direction: rtl;">
<span>(TTY/TDD) 711 </span>
</p>
```

```
<p style="direction: rtl;">
<span>&lrn; (TTY/TDD) 711</span>
</p>
```

```
<p style="direction: rtl;">
<span>&#8234; (TTY/TDD) 711&#8236;</span>
</p>
```

- As an alternative, LRE character can be inserted *before* the sequence, terminated by a PDF mark.
- To edit the XML (XLIFF) directly, use [Sublime](#), available on Windows, OS X and Linux as well. It works very well with Regular Expressions. It's a robust way to insert these marks using their unicode symbol, such as `‎`

Further readings:

- http://dotancohen.com/howto/rtl_right_to_left.html

3.3.4 SSL Certificates

Easyling has the ability to proxy HTTPS pages, but to do so, it must be provided a certificate and private key matching the URL. Otherwise, the proxy will be unable to identify itself as a valid server, and the browser will abort the connection for security reasons.

Easyling support can assist in deploying an HTTPS site by providing a CSR (Certificate Signing Request) to generate the appropriate certificate if the required information is provided, or the client can prepare the certificate themselves.

Additionally, a certificate is required to provide a branded Easyling instance. For more information, you can check the Whitelabel article of the FAQ.

The protocol

HTTPS (also called HTTP over TLS, HTTP over SSL, and HTTP Secure) is a protocol for secure communication over a computer network which is widely used on the Internet. HTTPS consists of communication over Hypertext Transfer Protocol (HTTP) within a connection encrypted by Transport Layer Security or its predecessor, Secure Sockets Layer. The main motivation for HTTPS is authentication of the visited website and protection of the privacy and integrity of the exchanged data.

In its popular deployment on the internet, HTTPS provides authentication of the website and associated web server with which one is communicating, which protects against man-in-the-middle attacks. Additionally, it provides bidirectional encryption of communications between a client and server, which protects against eavesdropping and tampering with and/or forging the contents of the communication. In practice, this provides a reasonable guarantee that one is communicating with precisely the website that one intended to communicate with (as opposed to an impostor), as well as ensuring that the contents of communications between the user and site cannot be read or forged by any third party.

Historically, HTTPS connections were primarily used for payment transactions on the World Wide Web, e-mail and for sensitive transactions in corporate information systems. In the late 2000s and early 2010s, HTTPS began to see widespread use for protecting page authenticity on all types of websites, securing accounts and keeping user communications, identity and web browsing private. (Courtesy of Wikipedia)

Issuing an SSL certificate

Issuing universally accepted certificates is restricted to the so-called “Root Certificate Authorities”. However, root authorities will generally delegate their powers to “Intermediate Authorities”, who will sign certificates as requested by the end user (provided ownership of the domain can be verified). When requesting a certificate for your white-label installation or HTTPS site, you will most likely interact with an intermediate authority, by providing them an encrypted configuration file (the Certificate Signing Request), which the provider consumes to produce a cryptographically signed certificate.

Using a CSR has several distinct advantages over generating your own certificate at the issuer: since the Easyling support crew is in control of the final product, we can tailor the request to generate the certificate we need from you; and since the private key remains safe with us, you do not need to take special precautions when sending it. However, we require some information to be provided to us beforehand:

1. `countryName_default = ${COUNTRY}`
2. `localityName_default = ${CITY}`
3. `streetAddress_default = ${ADDRESS}`
4. `postalCode_default = ${ZIP}`
5. `0.organizationName_default = ${COMPANY_NAME}`
6. `organizationalUnitName_default = ${ORG_UNIT}`

Issuing a certificate consists of several predetermined steps. First, a cryptographic key pair is generated, one half public, the other private. Then, a file indicating the domain the certificate is to be issued for, as well as various data about the entity holding the domain (generally, you or your client) is created. This file is then combined with the public half of the key, and signed with the private half. The resulting encrypted file is then handed to the issuer, who verifies the information contained within, and if successful, encodes the information into a public certificate, signing it with its own private key. This is the file that needs to be provided to Easyling support.

Providing us with a certificate

Following the previous phase, you are now in possession of a cryptographic certificate, and possibly its private key (if you elected to create your own certificate instead of requesting a CSR from us).

If we provided you with a CSR file, you need to send only the certificate. On its own, the certificate is not viable - it requires the private key to be useful, thus, it can be sent via email safely. The private key remains safe with us, and we will use it to upload the certificate to Google AppEngine, from where it will be available to authenticate the proxied site to the browser, and enable HTTPS for the translations.

If you elected to forgo the CSR and generate your own certificate, we will need the associated private key as well (and any passwords used to lock the private key). In this case, however, care must be taken to prevent the certificate falling into the wrong hands: the certificate and its keyfile, or the keyfile and its password must never travel together - if the email carrying both is intercepted, the malicious third party can use it to impersonate your site! Either use two separate emails, or even two different channels (email and Skype, for instance) to provide us the key and password and the certificate. Once we receive the files, we will again upload them to AppEngine after decrypting the keyfile, at which point it will be available for use with the proxied site.

N.B.: When selecting an SSL provider, bear in mind that Google only accepts certificates that are either self-signed or signed by a publicly trusted root Certificate Authority. One important point that needs to be highlighted is that the root CA for CloudFlare's Origin certificate system is *not* publicly trusted. Thus, we are unable to make use of Origin certificates generated via CloudFlare's system - in this case, please contact us for a CSR file for use with a provider of your choice.

SSL Manipulation Commands

Converting private keys to RSA

When uploading the keys into AppEngine, the file must be in RSA format. To verify, the beginning of the file should be

```
-----BEGIN RSA PRIVATE KEY-----.
```

If you read

```
-----BEGIN PRIVATE KEY-----,
```

you need to convert it, using the following command:

```
openssl rsa -in key -out key.rsa.key
```

Extracting Certificate and Private Key Files from a .pfx / PKCS#12 File (includes both the certificate and the private key)

- export the private key: `openssl pkcs12 -in certname.pfx -nocerts -out key.pem -nodes`
- export the certificate: `openssl pkcs12 -in certname.pfx -nokeys -out cert.pem`
- create RSA key / remove passphrase from the key: `openssl rsa -in key.pem -out server.key`

Check if a given key matches a certificate (or CSR)

By running these commands on the keyfile and certificate, you can verify that the key used to generate the certificate matches the one you have on hand. If the two outputs match, so do the keyfiles. But if not, your client may have used their own private keys to create the certificate, which you will have to obtain before forwarding it to us.

- `openssl rsa -noout -modulus -in privateKey.key | openssl md5`
- `openssl x509 -noout -modulus -in certificate.crt | openssl md5`

Alternatively, if you have the CSR as well, you can use the following command to obtain the checksum of the CSR's key and verify that the CSR you have on hand was used to generate the certificate.

- `openssl req -noout -modulus -in CSR.csr | openssl md5`

3.3.5 Proxy modes - X, P, live

X-proxy (testing, JS bugs, JS fix domains , CORS)

The X-proxy is great for testing. You can spot content that does not get picked up by default, and make your configurations to your project, and check for success.

There are a couple of situations, when the X-proxy comes in handy:

- Testing regular expressions, for example on e-commerce sites.
- Testing JSON (JavaScript) and XML translation.
- Just browsing through a site, for evaluation purposes.

An example X-proxy URL: `https://de-de-{project_code}-x.app.easyling.com`

The X-proxy can be accessed from the pages list under Content (or Discovery) by clicking on the Preview button in the hover toolbar, while holding down the Ctrl/Cmd button, or you can just replace the -p for a -x in the normal preview's URL for the same effect.

P - Preview

The standard proxy mode to view the translated website before publishing. However, the preview can be used for a couple of other things:

- Cookie header extraction to get behind logins
- Visiting pages manually, to ingest content

An example Preview-proxy URL: `https://de-de-{project_code}-p.app.easyling.com`

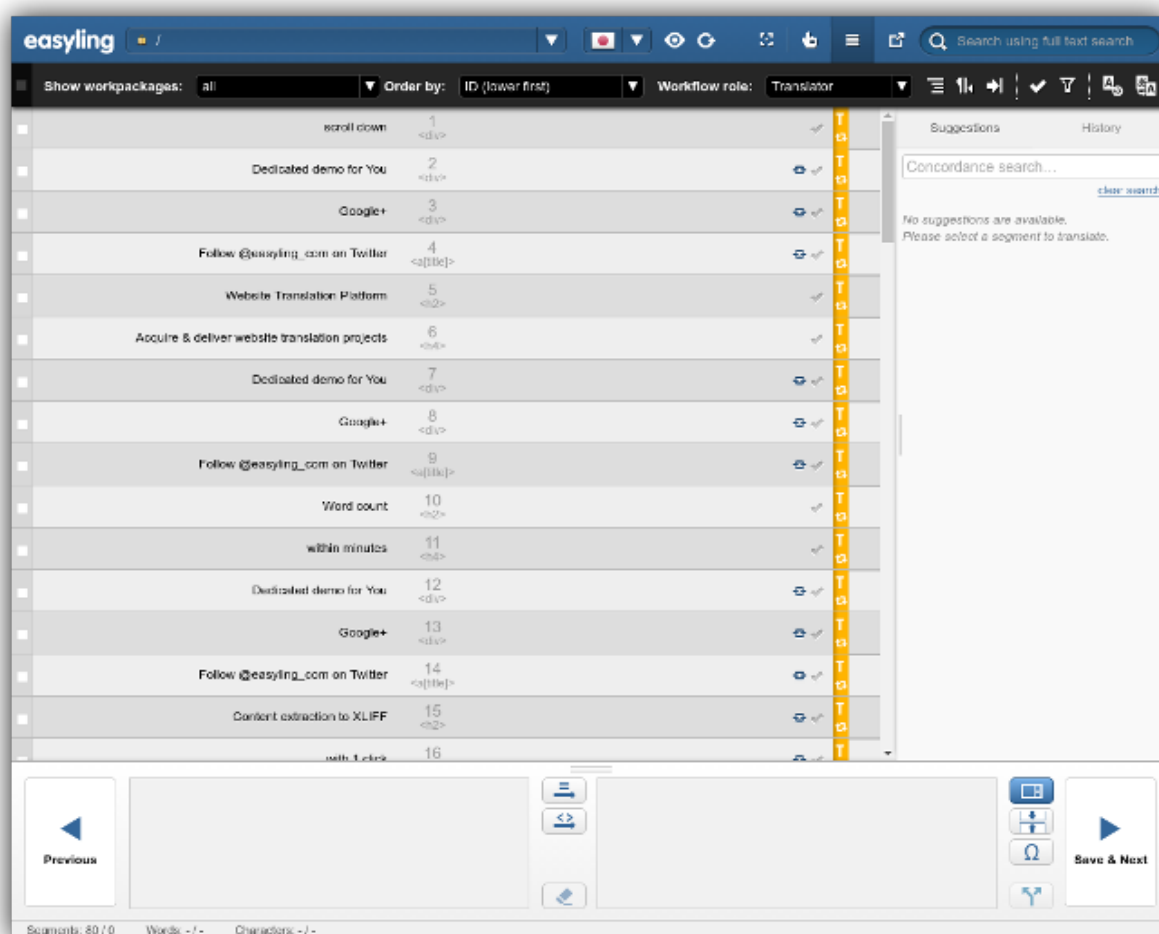
C - Channel

Live serving mode

After publishing the website, the proxy serves content on the chosen domain.

The Workbench

4.1 Introduction



The Workbench is a CAT tool in the cloud.

If the Dashboard is our take on the project management side of website localization, then the Workbench is the analogous CAT tool you can use without ever leaving the browser window.

In this chapter, we take a close look at all the various features available on the Workbench. The explanations pertaining to the various elements and functionality of the Workbench screen are grouped into thematically linked subsections. Use the sidebar to navigate to the subsection of your choice.

4.1.1 Opening the Workbench

You can access the Workbench anytime after adding at least one target language.

There are a few ways to enter it:

1. In the Dashboard menu, see the *Languages* part. If you hover over any of the target languages, a menu bar will be displayed to the right. Click on “Translate in List View”..
2. Click on “Manage segments” in Content menu...
3. Enter the Page list and hover over any entry to display a menu to the right. Click on “Translate in List View”...

.. and the Workbench will open in a new tab.

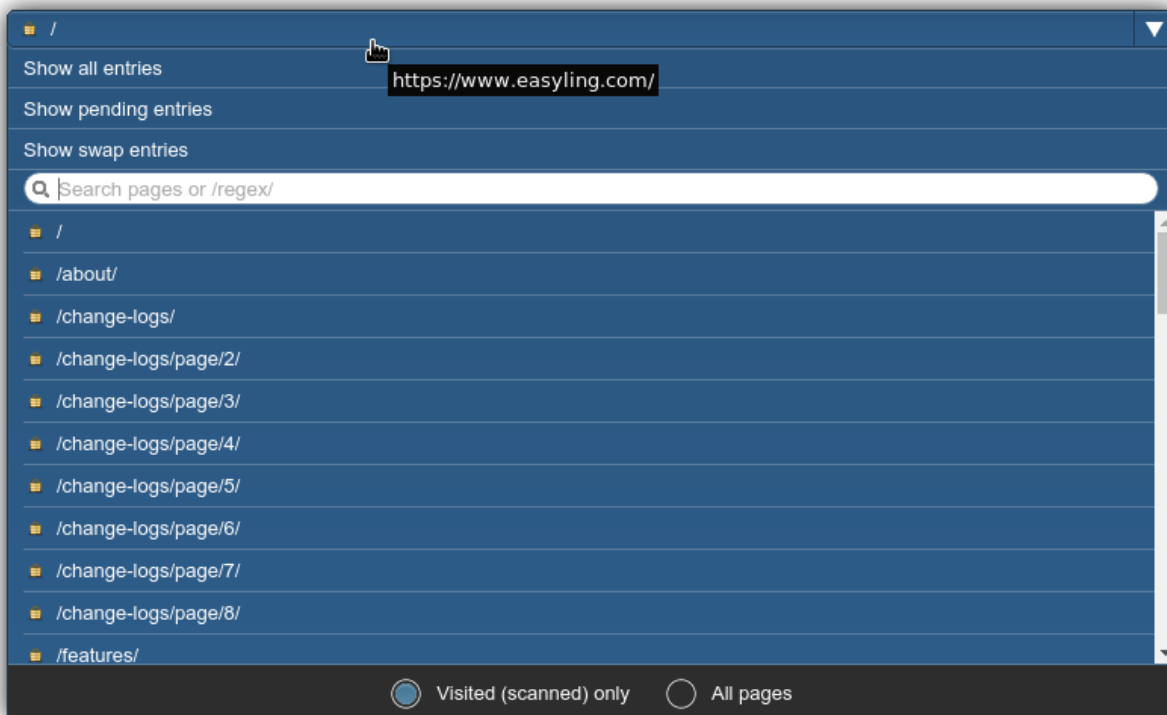
4.2 Moving Around

The Workbench has a single viewport, so every feature you’d need to navigate between pages and segments is always just a couple of clicks away.

You may switch between pages, search for text, filter for segments based on a variety of metadata, such as Approval state, containing block element, translation source and so forth. This section deals with the various ways to do that.

4.2.1 Workbench Page List

A segment is tied to the specific page it was found on. You can use the Page list dropdown right next to the logo to get a list of all pages currently within the scope of translation.



You can click on any page entry in the list to visit that page and get an overview of all segments associated with that page. Websites can get very large with a huge list of pages - use the search field embedded in this dropdown to locate a specific page (you may use regular expressions with the usual format of including them between two slashes like this: `/[regular expression]/`).

There are three options at the top of the dropdown that bear special mention:

Show All Entries

Most of the time, you will want segments displayed for a specific page, but you may also use this option to get an overview of all segments across all pages.

WARNING! Only **List View** is available in this view, all other View buttons will be unavailable!

The All Entries list doesn't flood your browser with every last segment all at once: segments will be loaded in batches of 500. The Workbench will automatically fetch a new batch of entries as you scroll down.

Show Pending Entries

By default, Scans will pick up new entries in "Approved" state, which in this case means "Approved for Translation", immediately available for translation. You can change this default behavior, by going into the Dashboard and changing it in Advanced settings to either Pending or Excluded.

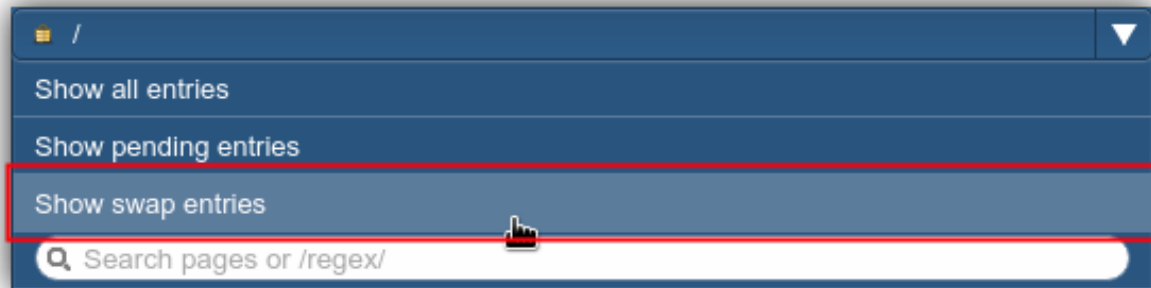
By clicking on "Show pending entries" in the page list, the list of all entries that are currently waiting for Approval is displayed. In their current state, They will not be included into exports unless the relevant option is selected at export-time, and will not appear for translation unless filtered for specifically.

Project or backup owners, or users with the Customer role can move these into either one of the other two states, by **approving** them or **excluding** them entirely from the scope of translation.

Show Swap Entries

Swap entries are those segments that have had the “EL_swap” class added to their enclosing tags on the source site.

They are special in that they are added to the Workbench without processing their tags. They will be displayed verbatim, allowing you to edit the source content markup directly. That content will be sent as-is by the proxy for each request.



Take caution when editing swap entries. All responsibility of rendering them successfully *and* safely is delegated to the requesting browser.

Filters

There is a comprehensive assortment of filters available. Click on the Filters icon in the toolbar to get a an overview of all available filters:

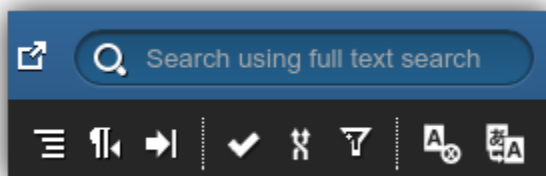


Use the checkboxes to define your filtering settings in the dialog and click on “Set Filters”.

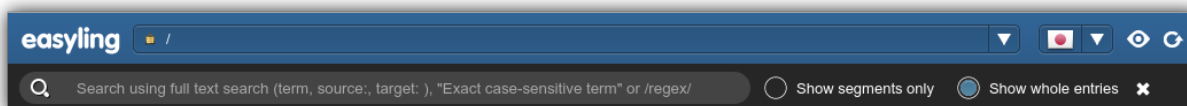


These filters work with the various types of metadata associated with an Entry (such as currently assigned workflow role, enclosing block element in source, approval state or source of translation), **not** content - use the **Search** functionality to filter for source or target text.

4.2.2 Searching for segments



Use the full-text Search field in the upper-right hand corner of the screen to search for segments. When you click on the Search field, the Workflow part of the black menu bar will be superimposed with a search field right above the Viewport.



Normal full-text search and regular expressions are supported. The normal search can be a bit misleading, as it is a forward-from-word-boundary search. If, for example, you'd like to find "translation" in a source segment, a search query for "translatio" will turn up all segments containing any words that start with that exact set of strings (and likely end with n,ns, or nese, perhaps). Searching for "ranslation" only, however, will return zero matches.

Use regular expressions to search between word boundaries: to extend the previous example, a query for `/.ranslation/` will show all segments that contain any character (except newline) exactly once, followed by the literal string `ranslation`. The same as with the Page list, a string is interpreted as a regular expression if you enclose it between slashes.

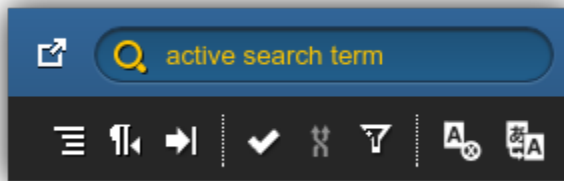
You may select between displaying segments or whole entries using the radio buttons next to the Search field.

Closing the Search Field

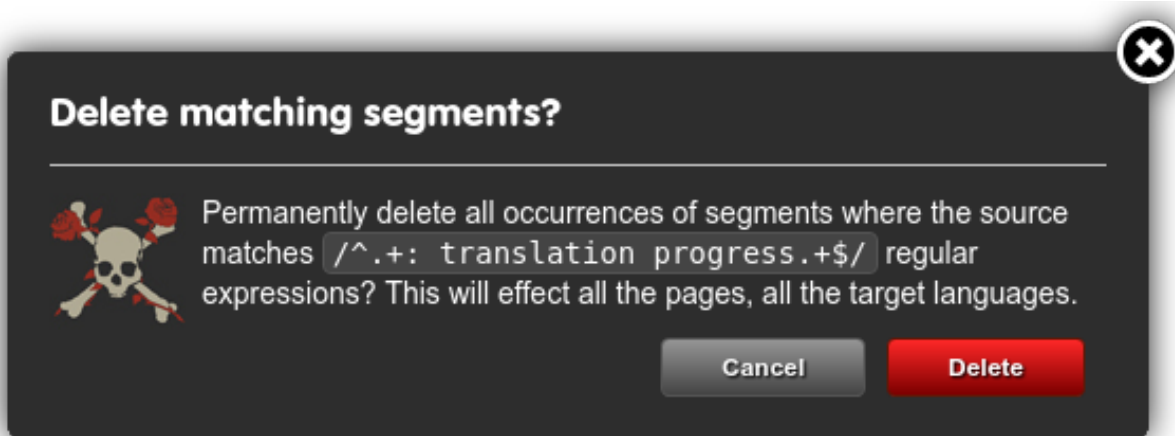
It is important to remember that the search field is also a **filter**: as long as it is active, segments will be filtered based on its contents regardless of being in All entries view or on a specific page.

If you wish to restore full view of segments, clear the search field and send an empty search. Use the close button next to the Search display options to close the search field.

If a search string is still present, it will be preserved and displayed in the upper right corner Search field in orange, like this:



Deletion of Segments



You might be wondering what this has to do with navigation, but executing a regex search reveals another feature of the Workbench you might ordinarily look for someplace else: that of deleting segments. If you click on the Magnifying glass icon while the regex search is active (a bottom-facing triangle will indicate availability), the dialog above will open.

The non-discoverability of this feature is premeditated. Skulls & Bones warnings would generally apply to **any** situation where the words “delete” and “regex” are found in the same sentence. That being said, to give you a measure of peace of mind, deletion of segments is not as final as we seem to make it out to be: when TM Freeze is **disabled**, you can re-add segments anytime by Scanning or visiting in Preview the page that contains the deleted segments.

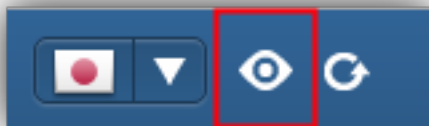
But that is new words added each time. So, if for no other reason, be careful about deleting segments in order to avoid unnecessary expenses. Buyer, beware!



4.2.3 Preview

By clicking on the “Eye” icon, you can visit the temporary domain to check your translations in their original context. In the “All Entries” view, this function is disabled if no segment is selected - without picking a segment, the Workbench has no information on which page to show you. Otherwise, if a Page view is open, the selected page will be loaded.

If you select a segment in All Entries view, the Preview proxy will open on the page where it was seen by the crawler for the first time.

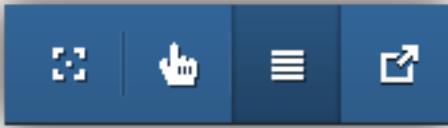


The icon on the Workbench will take you to the Preview mode, but there are a few different Proxy modes available besides that. See [here](#) for details.

There is more than one way of looking at content on the Workbench, and the default, the List View is only one of them. In this section, we go over the various in-context ‘Views’ you can access from the Workbench and use to edit your translation while making sure that it is behaving exactly as it should in the original context.

4.3 Views

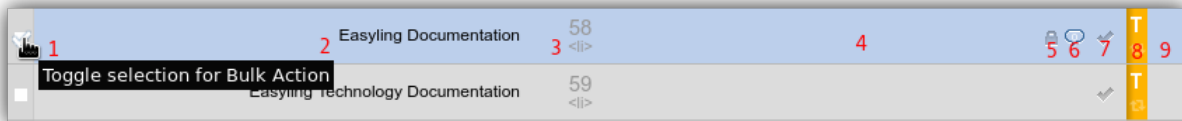
4.3.1 List View



While not necessarily the most impressive, the List View is certainly one of the most useful views on the Workbench. You can use it to go over each segment being translated and edit, filter and search for any subset.

The List view provides various features to use with each entry. The currently selected Entry will be highlighted in yellow, hovering will causes the Entry under the cursor to highlight in blue. The presentation of Entries is clear and simple, yet there are a variety of features you can use with each. Let's take a detailed look at an entry and see what each part of a line does.

Anatomy of an Entry

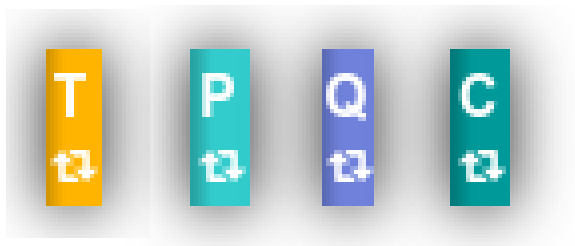


- Select Checkbox:** check this box to select the entry or segment in question. You may use the “Bulk Actions” icon to batch process selected entries (i.e. confirm, exclude or approve them)
- Source Entry:** contains the text that was found within a given block element on the source site. You may set the text direction using the “Align source text” icon in the toolbar. Otherwise, the Source Entry is not editable.
- Entry No. + Containing Block Element:** use the “Go to Segment” icon to jump to a segment with a given number. This number is assigned to entries in order of arrival (new entries can be found at the bottom of the list). Additionally, the tag that contained the text on the original site is displayed below it (useful when you want to identify a segment in the HTML source).
- Target entry:** The translation, provided by a variety of sources: Manual Editing, Translation Imports, Machine Translation or Translation Memory.
- Lock Segment:** Prevent any changes from influencing the current content of the Target Entry. Especially useful when you want to run batch processes on your segments, but you want to exclude an entry from the scope.
- Comment on Segment:** Use this icon to add comments to an Entry either as a note-to-self, or as part of a collaborative translation effort. All comments have a checkbox next to them, allowing you to mark them as settled.
- Chain link:** indicates that the segment is repeated verbatim (102% match) in the current view. Click on the icon to jump to the next repetition.
- Confirm Tick** click on the tick to Confirm the segment for the current workflow and send it to the next stage. Confirmed entries remain editable as long as they are unedited by the next workflow role.

9. **Workflow Status Indicator:** Display the current workflow state of the segment. Note that this might differ depending on which user in which workflow role is currently looking at the segment.
10. **Flag:** Display current translation source.

Workflow Tags

Depending on what Workflow role you're currently in, the following Workflow tags will be displayed for each segment, influencing their availability for editing:



T - Translator

P - Proofreader

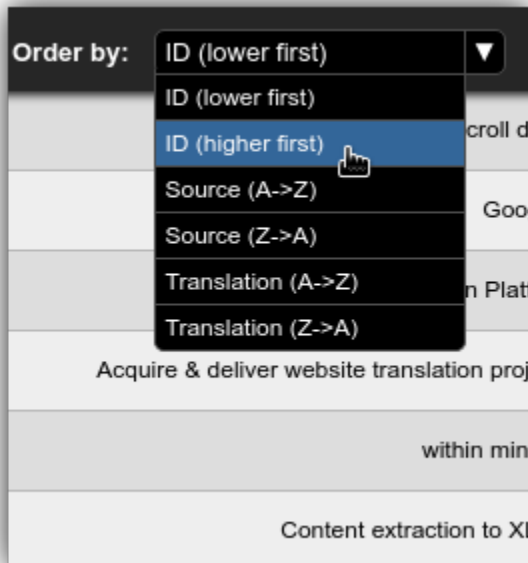
Q - Proofreader 2 (Quality Check)

C - Customer (or Client)

See the section on Workflow roles for a detailed description of the various workflow roles.

Flags

Ordering Segments in List View



Use the “Order by” dropdown to alphabetize the target or source entries or reverse the order of segments based on their Workbench ID. The following Ordering methods are available from the dropdown:

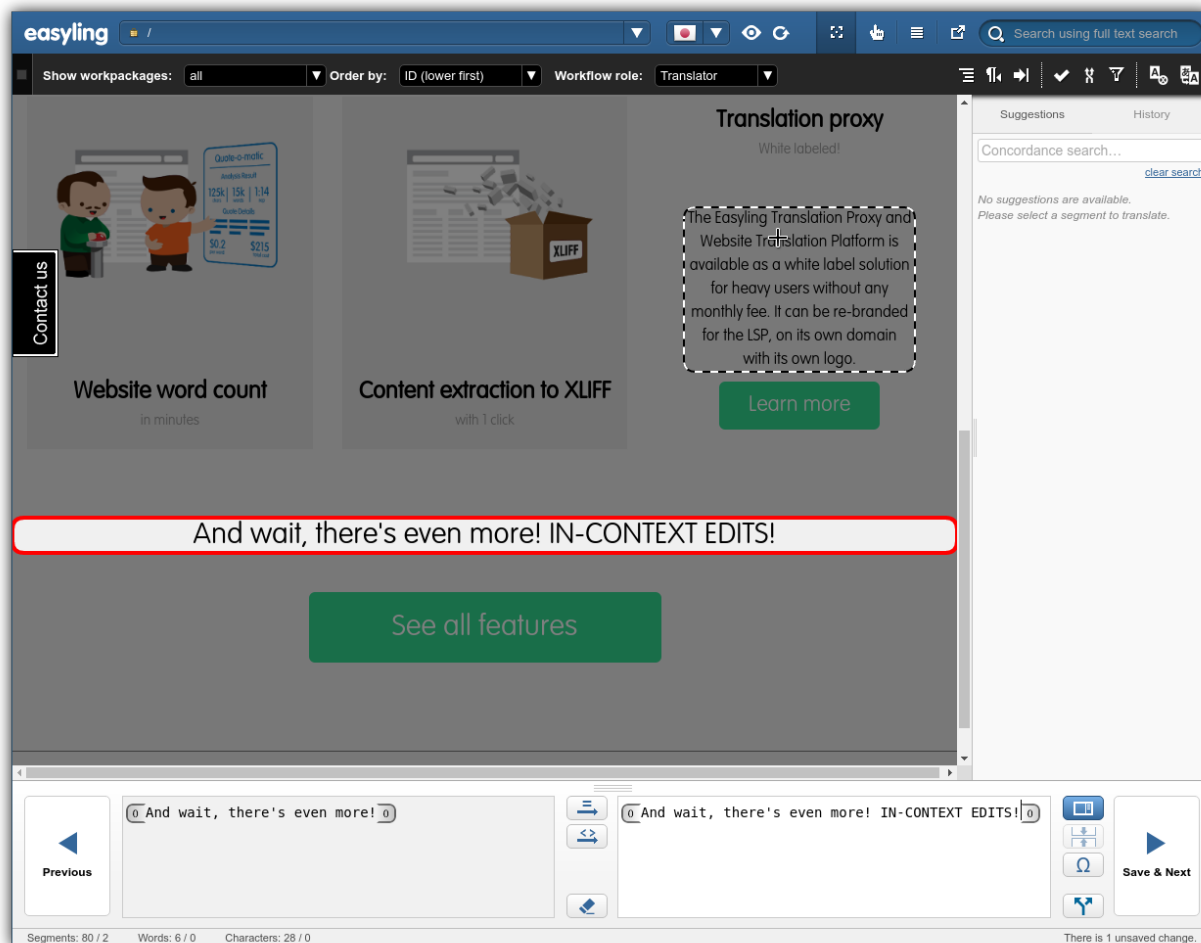
- **ID, lower first** (default)
- **ID, higher first** (use to quickly navigate to new segments)
- **Source A → Z**
- **Source Z → A**
- **Translation A → Z**
- **Translation Z → A**

4.3.2 Highlight View



The highlight view is a true in-context editing view that makes the Workbench popular with Translators, and *the* solution to the problem of adequate context during website localization.

By selecting an Entry in List view and clicking on the Highlight View on the Workbench, you will be shown the text on the original webpage



You may click on any part of a website and have a highlighting frame appear around that segment. At the same time, the editing box below will jump to the segment in question, where you can add/edit your translations in-place.

Really, the Highlight View is simplicity itself with very little in the way of hidden gotchas. Select a page, point & click, and translate away!

But keep in mind that while you are using the Highlight View with a page, links will be unclickable - use Free-click View to navigate on the original site from within the Workbench.

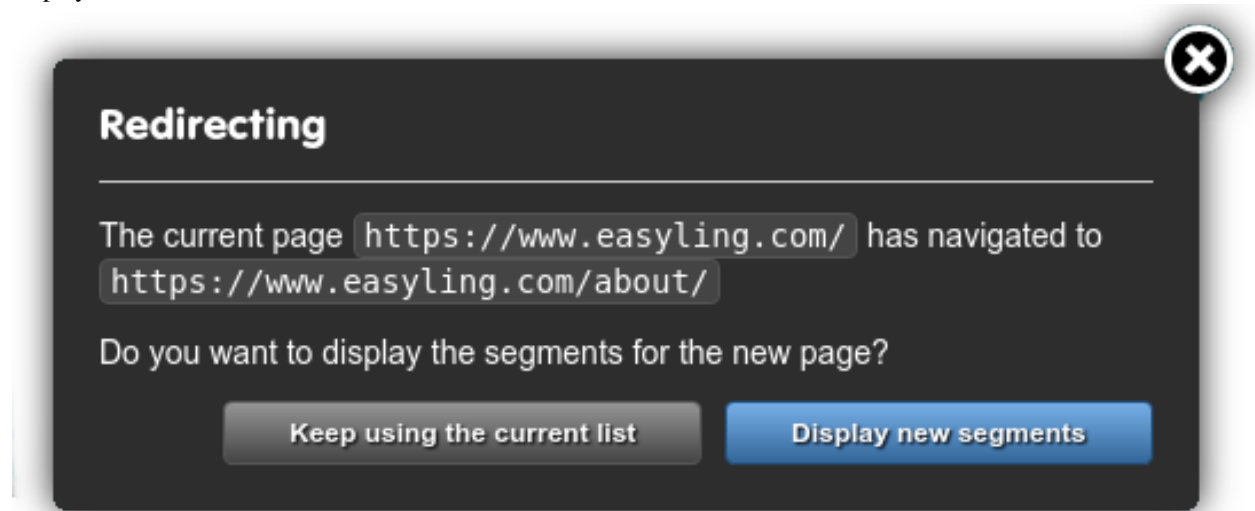
NOTE The Highlight view is a wonderful tool we are very proud of, but don't forget that much of the textual content of a website is not clickable. Check the other modes and the Preview to make sure that everything is covered!

4.3.3 Free-Click View



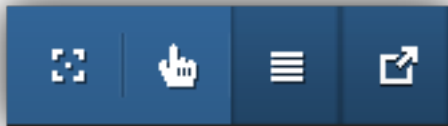
The Free-Click View is much like the Highlight View, but it allows you to use the links on a site to visit other parts of the website. Alternate between Free-click view and Highlight View to translate segments as you explore the website.

Free-click view will offer to reload your segment list based on which page you are on. The following dialog will be displayed.



Highlight View will only work if the segment list is that of the current page.

4.3.4 Pop-out View



The Workbench is a single-viewport application, and the Pop-out View is a feature meant to allow you to have the Highlight View and the List View on your screen **at the same time**. Select an Entry from the list and click on the Pop-out view icon to open a Highlight View in a separate pop-up window.

Modern browsers block popup windows by default, so you will most likely have to **enable popups** for the Workbench for this feature to work.

4.4 Translation of Segments

We are getting to the primary purpose of the Workbench, which is of course, translation of text. In this section, we'll talk about the different ways of doing that.

It bears mentioning right off the bat that the Workbench is not meant to selfishly yank you out of your CAT tool of choice. As the proxy supports exporting segments in the industry-standard XLIFF format, you can always employ any tool of your preference, SDL Trados and MemoQ, for example, being two major players in the field. The choice is yours.

Nevertheless, you will find that the Workbench is closer to where the website really is - up there in the cloud, and therefore immensely useful during the editing process. You'll see it truly begins to shine when your XLIFF files have done their first round-trip to and from external CAT tools.

There are multiple ways of translating content on the Workbench:

- manual Translation using the Target editor
- using Translation Memories
- using Machine Translation
- importing translated XLIFF files

The following sections provide an overview of the translation methods listed above.

4.4.1 Editing segments

Editing of segments happens in the Target editor located on the bottom of the Workbench:



You can use the Highlight View or the List View to select a segment, and the editor will reflect your selection. The Source entry is displayed to the left, the translation (Target) to the right. Only the target is editable.

Editing Window

Editing translations on the Web means going around/avoiding/leaving intact the forest of HTML tags that the text is usually embedded on. The Workbench abstracts away these markup details to ease working with text (the first and foremost task of translators), but doesn't, and will never attempt to hide the fact of their existence.

Tags are represented as numbered grey widgets around certain parts of the text, which you can use as a yardstick to place your translations in the appropriate tag context without having to worry about what those tags actually do.

Two things:

- **You can Drag & Drop the Numbered Tags**
- **You can NOT delete them**

Adding translations is otherwise straightforward text input. Untranslated entries will contain the Source text as a placeholder until such time as a change is made to the contents of the Target, at which point both the List View and the Segment contents are updated.

There are a few smaller buttons in various parts of the Target editor. In the middle, you'll see these three buttons:



These contain editing functions that relate the contents of the Source and the Target in different ways. Use the top “**Equal**” button to copy the Source contents to the Target, rendering them identical. You can use the middle, “**Tag**” button to copy the tag structure of the Source to the Target. The lower “**Eraser**” button will delete the translation and restore the placeholder.

There is another set of four buttons to the right:



They all deal with the Target. In descending order, these are:

1. **Toggle sidebar**: use this to hide the Suggestions & History sidebar above the button.
2. **Preserve Whitespace**: prevent trimming of whitespaces in the segment
3. **Insert non-printing characters** There are a number of non-printing characters you might need during a translation project, especially if you are dealing with languages that have a Right-to-Left writing direction, such as Arabic. See the section on RTL conversion for further details.
4. **Split segment from group**

Saving changes

Click on “Save & Next” to save your work on the given Entry and jump to the next segment. You can also use Ctrl+Up or Ctrl+Down to do this. This is an explicit, although not strictly necessary action, as any edits are **automatically**

saved upon leaving a page, or otherwise after 60 seconds of inactivity on the Workbench. Any navigation generated by the Workbench itself will also trigger a flush of all unsaved segments.

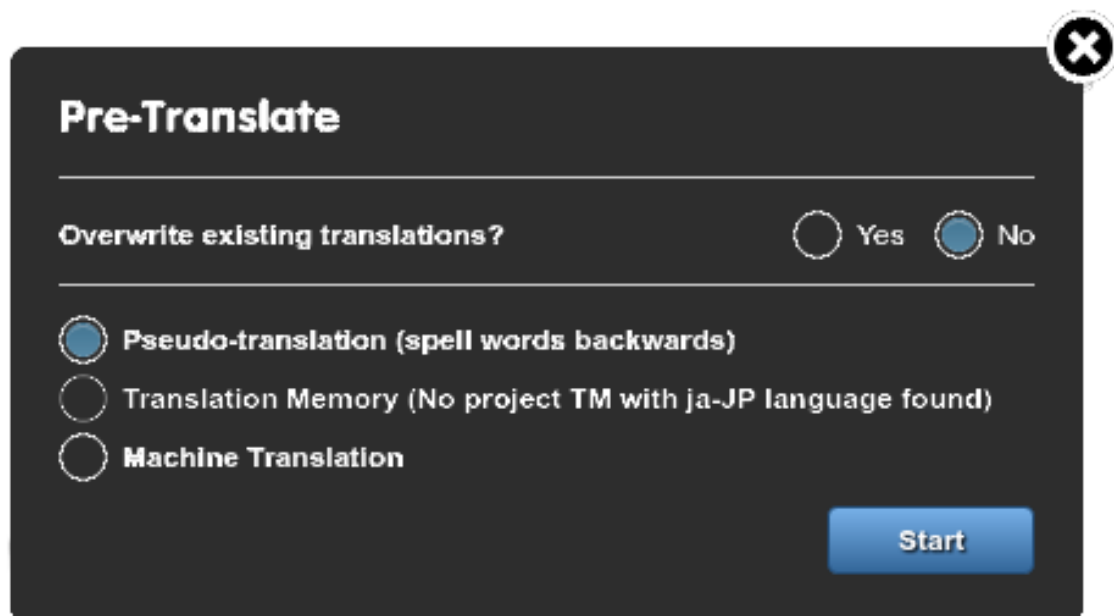
4.4.2 Automatic Translation

There are a number of Auto-translation features that you can employ to ensure quality on the proxy.

The option to run pre-translation or set it up to run automatically is also available on the Dashboard. Here, we discuss the options available after clicking on the “Pre-translate” icon in the toolbar:



The following dialog is displayed:



Pseudo-translation

This function is admittedly not translation in a meaningful sense, but string-reversing each and every word of every segment on the spot is very effective during a demo. Use it, then go to Preview to demonstrate to the client that website localization is as painless as using CAT tools in tandem with the Proxy. Good for the wow!-factor. You can also use pseudo-translation to test the various editing features available on the Dashboard and the Workbench.

Translation Memory

If you have a populated Translation Memory on a project, you may use its contents to translate segments. Use this feature to translate your content with a preset match threshold.

Machine Translation

You can choose to Machine Translate the currently selected batch of entries using one of the available MT APIs (Google Translate, Bing Translate, iTranslate4u and GeoFluent).

Translation Memory and Machine pre-translation are both reproductions of the options accessible on the Dashboard, with the added functionality of being able to control which specific group of segments should be targeted by the process.

4.4.3 Search & Replace

The Workbench has a Search & Replace feature you can access by clicking on this icon in List View:



It always operates on the currently listed set (or subset) of editable segments. The dialog that opens will look like this:

Search and Replace

This Search & Replace function operates on the set of editable segments currently listed, and the changes are saved in the database (not in TM). Once the operation is ready, you receive an e-mail notification with a detailed report on the entries affected. **Please note that the operation cannot be undone!**

Details

Replacement scope: <https://www.easyling.com/about/> Target locale: Workpackage: all Affected entry count: 78

Filter by source ☐ Regex

Search in target ☐ Regex

Replace in target

Preview

Preview unavailable

To preview the results of the replacement, tick in 'preview changes' and perform a search

☒ Preview changes on a few segments first ☒ Test run (no changes will be made)

Both ‘Search in target’ and ‘Replace in target’ values are required, while ‘Filter by source’ is optional. You can use either simple strings or regular expressions as search terms (simple strings are the default, check the ‘Regex’ option next to the field to use a regular expression). The function is case-sensitive.

Choose ‘Test run (no changes will be made)’ to check how the operation would affect your translation. To see the entries affected by replacement, check the ‘Preview changes on a few segments first’; the segments will appear in the Preview area (Preview can’t be used without Test run). You can also start replacement right away, without preliminary checking.

Once the operation is ready, you will receive an e-mail notification with a detailed report on the entries affected.

Please note that the replacement is done in the database, not in the TM. If you want your TM to reflect these changes, you need to run ‘Populate TM’ from the Dashboard.

Keep in mind that **the operation cannot be undone!**

4.4.4 History

The proxy keeps tabs on what happens to each Entry in the project timeline, and the Workbench displays these tabs in the sidebar. You can use these to access previous editing states of an Entry. Here follows a short description of the History functionality.

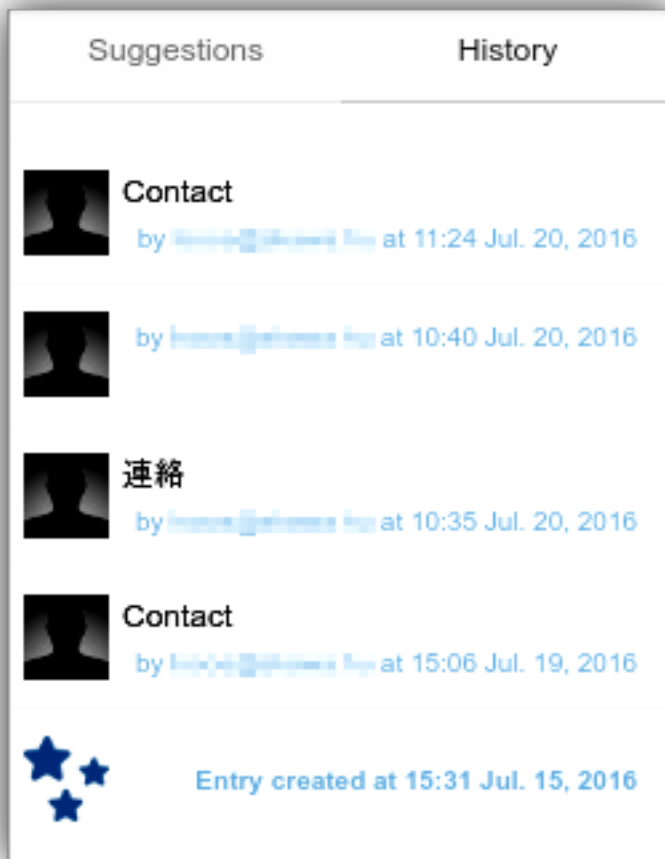
Translation Memories

If you select a segment, TM suggestions will be displayed for it in the sidebar tab labeled “Suggestions”. Click on any one of them to add it as a translation for the given Entry.

A Search field is also provided that you can use for concordance lookups.

Segment History

Whether a result of a manual or an automatic edit, each saved state of an Entry will be saved with a username and a timestamp in the Entry history. You can access it in the sidebar tab labeled “History”.



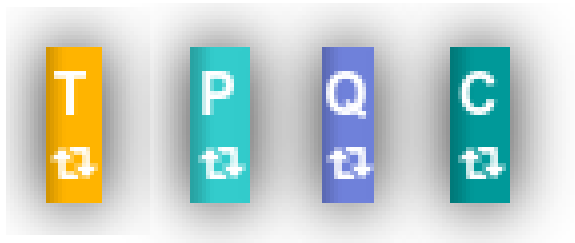
This means you don't have to worry about ever losing translated content as a result of manual edits - you can always restore a previous state of an Entry by selecting the Entry in List View or Highlight View, going to History and copy & pasting a previous state of your choosing.

4.5 Collaboration

Collaboration are a must with website localization. If you have used the Dashboard's Sharing settings to invite other people into the project, granted them the appropriate editing rights, the list of segments in the Workbench will look a bit different to each user.

4.5.1 Workflow Roles

As mentioned previously, roles are predominantly a project management feature associated with work on the Workbench. To reiterate, there are four different roles:



T - Translator (default)

P - Proofreader

Q - Proofreader 2 (Quality Check)

C - Customer (or Client)

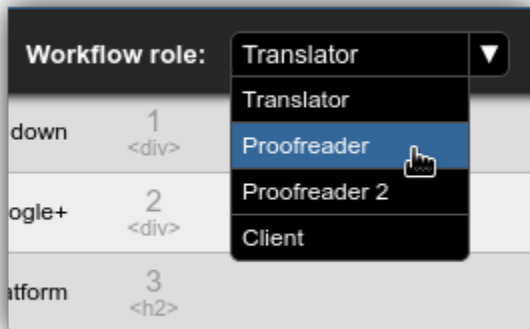
There are four different workflows on the proxy you may employ. You may set these on the Dashboard.

- Simple Translation Workflow (**T**)
- Translation + Proofreading (**TP**)
- Translation + Proofreading + Client Approval (**TPC**)
- Translation + Proofreading + Quality Check + Client Approval (**TPQC**)

Each setting will activate the necessary roles, which the Owner or Backup Owner may assign to any project participant. By default, only the Translator role is required. Owners have access to all workflow roles.

Workflow Roles in Action

Use the Workflow role dropdown in the toolbar to switch between the available Workflow roles:



Take TPQC, the workflow with the most participants, for example.

1. Each **approved** segment is assigned to the Translator role.
2. When finished with a segment (either through manual edits, automatic translation or via XLIFF importing), the translator clicks on the Confirm tick to declare that segment cleared for that phase and send it to the next role, the *proofreader*.

3. The *proofreader* (and much everyone else) may use Filters to display only those segments that are assigned to that role. He takes the segments sent by the Translator, edits them, changes their wording as required. When finished with an entry, the *proofreader* clicks on the Confirm tick, sending the segment along to *Quality Check*.

And so on. This cycle is then repeated until a segment (more to the point, *all* segments) reach the final workflow role, that of the Customer, who approves translated entries.

A few things to keep in mind:

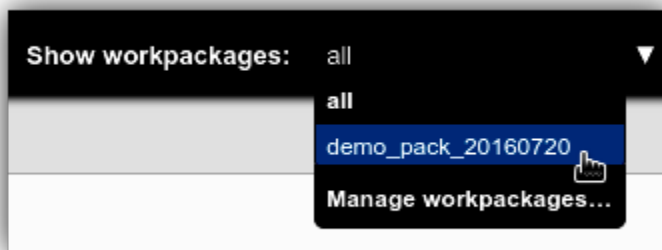
- Each Role has access to the lists of **upstream** roles.
- Only Owners, Backup Owners and Project Managers have access to **all** roles.
- Entries/segments belonging to another role are **greyed out**.
- A segment remains available for editing after Confirming it just as long as it is not touched by the next Workflow role. If you ever mistakenly Confirm an entry, you may, so to speak, **reclaim** it for some more work before the next role can get to it.

And that's about it!

4.5.2 Work Packages

If Workflow Roles is a method of grouping your users, then Work Packages are a method of grouping segments. See the Dashboard chapter for the details of how to generate them.

Use the Work Package dropdown to select a Work package, and the Workbench will display only those Entries that belong to that Work Package (note that an Entry may belong to more than one Work Package!). The dropdown looks like this.



The only default entry in the dropdown is “All”, which means disabling Work Package based filtering on the Workbench. As new Work packages are generated, this list will automatically update after refreshing the window. The dropdown will always contain the names of the 100 latest Work Packages.

Clicking on “Manage workpackages...” will take you to the Dashboard where you can tend to your existing Work Packages or generate new ones.

4.6 Cookbook

4.6.1 Search & Replace

Tutorial 1: Fix Spelling (String Replacement)

A simple use case of the *search & replace* feature is the old chestnut: the differences between British and American spelling rules.

This could come up whenever you are working on both the en-US and en-GB locales, or if two translators, each on a different side of the pond, forgot to coordinate their spelling.

Let's say you have the following targets with German as a source language:

```
The world's No.1 donut
Vanilla donut
Chocolate-chip donut
Doughnut miss it!
```

and so on. Replace all instances of the word `donut` with the `doughnut` variant by following the steps below.

1. Click on the Search & Replace icon



1. Fill in the `Filter by source` field to work exclusively on those entries that contain the given string in the source language. This example is about the spelling of “donut”, so you would enter the German original, `Krapfen`, to limit the search. Rest assured: source entries are **never** changed.
2. Enter the word that you'd like to replace in the `Search in target` field, in this case, `donut`.
3. Enter the replacement, `doughnut`, in the `Replace in target` field.

That's all there is to setting things up, the rest is about making sure your changes will not cause any problems.

1. Click on **Preview!** to see your changes applied to a subset of segments.

2. If contents of the Preview area look good, uncheck the Preview checkbox.
3. (*Optionally*)do a Test run by clicking on “Go!”
4. Uncheck the Test run checkbox and click on “Go!” to really apply the replacement.

Depending on the number of segments, the process can take some time to finish. You will receive an e-mail for both the Test and Live modes, containing a list of proposed (Test) or applied (Live) changes.

Tutorial 2: to-do (Regex Replace)

Maintenance

5.1 Project maintenance

The translation project of a website is a continuous task, as new content is regularly added to the site. So uploading the initial translation is not the end of the project, but the beginning of a new phase. This phase is practically a repetitive cycle of the following activities, many of which can be automated:

- checking the site for new content
- extracting new content for translation
- translation of the new content
- uploading the new translation
- and, of course, error fixing, if needed

5.2 Automation possibilities

The first two activities can be set to be carried out automatically at daily, weekly or monthly intervals, depending on how frequently content is updated on the website. This is called **scheduled crawl**. If you turn on this feature, changes will be checked and retrieved for translation at the specified intervals, and e-mail notification will be sent to all project participants. The new content is available right away for translation in the online editor interface, and you can also download an XLIFF for translation.

Please note that this check is technically content extraction, so it has an associated cost of EUR 2 per 1000 words.

You can enable this option in **Content > Settings > Look for changes**. Of course, you can set it to **Only manually**, if the site does not update frequently.

Website content

Discovery helped identifying pages relevant for translation. **Scan now:** Easyling will visit the selected pages, retrieve the content for translation and store it in the database. Subsequent **scanning** will also discover new pages and new/changed content on the site.

Settings

☐ Process pages in source language only (requires Google language detection API)

Ignore queries

Group pages by ignoring query parameters

Matching source segments are merged across pages in the group

Discovery cookies

These values are not saved for security reasons. [More about cookies](#)

Scan page limit number of pages maximum Leave blank for Unlimited

Look for changes or pages: 8, Time elapsed: 00:00:15.


Statistics

The process can automatically extract new content, apply the associated translation memories and machine translation services, prepare a work package and the XLIFF export. This is particularly useful for fast-moving sites where content arrives quickly and time spent untranslated needs to be minimized (possibly at the expense of real-time human oversight).

5.3 Automatic pre-translation

Easyling can automatically pre-translate incoming content without user intervention or oversight, feeding its translation engine from saved translations (from an Easyling Translation Memory) above a certain confidence level or using configured machine translation engines (Google Translate, Bing Translate, iTranslate4EU, and GeoFluent, currently).

If new content is encountered, and at least one source is configured, a user-configurable timer starts counting down. Content is collected during this timer, and automatically translated using the configured sources. At the end of the configured window, any content that cannot be translated with the assigned sources (no matches of the desired confidence are found in the TM or the MT-engine returns no translation) is packaged into an auto-generated Work Package and exported into an XLIFF file (being pushed directly to XTM, if an integration is configured). The resulting export can then be translated in any external system and imported back normally.



Auto pre-translate

In this dialog you can set, how to handle new entries when detected. (The auto pre-translate process will run if any of the following options are enabled.)

☐ 1. Translate with Translation Memory

☐ 2. Use Machine Translation

☒ 3. Generate Workpackages

Gather new entries for minutes before auto-generating Work Package.

☐ 4. Send XLIFF files to external system

Send XLIFF files to

[Configure XTM](#) [Configure XTRF](#) [Configure Dropbox](#) [Save](#)

5.4 Caching And The Proxy

5.4.1 Overview

Nearly all browsers today implement local caches to accelerate page loading and prevent unnecessary requests from being sent out to the network. However, the operation of these caches is tied to the presence of certain headers on the page, such as `Pragma` and `Cache-Control` - based on their presence and the values communicated in these headers, the browser (and various systems, such as CDNs) may make a decision to intercept the request and serve up certain content without requesting it anew from the server.

Normally, the Proxy simply forwards these headers, much the same way it does with any other. The option to override their presence and values exists (see the Path Settings option on the Dashboard), but by default, they are left unmodified, in the spirit of minimum invasion. This is not always desirable, however, as a site without such cache headers will remain uncached in the visitors' devices, and each visit to the page will result in another request that is billed.

5.4.2 Inspecting Cache Headers

You can investigate how well a site may be cached using the Developer Tools in most major browsers. In Chrome, for instance, the DevTool can be summoned using by pressing the F12 key (or Alt+Cmd+I on under Mac OS), and after refreshing the page, the Network tab can be used to browse traffic associated with the tab. By selecting any entry in the list, you can view its details, in particular, the request and response headers. To tell whether or not a given resource will be requested again, you need to look at the “Response Headers” section, and look for the keys `Cache-Control` and `Pragma`.

If you see `Pragma:private` and/or `Cache-control:no-cache`, it is safe to say that the given resource will not be cached and each visitor will result in another hit. Files like this will likely prove resource drains if the site receives large amounts of traffic. On the other hand, `Pragma:public` and `Cache-Control:public, max-age=\d+` (where `\d+` means at least one digit, or more) are good signs in that these files will be stored on the client’s device after the first request, and will not be requested until `max-age` seconds have elapsed since the last load, and will save resources in the long run. Of course, this also means that visitors may be seeing an “outdated” version of the resource for a limited time before their caches expire and are reloaded.

There is a bit of a gray area when seeing `Cache-Control:must-revalidate`: this directive allows the cache to make the final decision based on its own algorithm, and the response **may** be stored, but not necessarily. When seeing this it is good to prepare for potential increased traffic, as the browser cache may or may not retain these responses.

5.4.3 What to take away

If you’re experiencing consistently large numbers of page views on a given project, it is often a good idea to suspect caches (as opposed to search bots, which cause transient spikes). In such cases, you should inspect a few pages using the DevTool, and determine whether or not the site is set up to take advantage of browser caches.

If it turns out be the case that the site is not set up to use caches, the best course of action is to notify the owners - perhaps they have a good reason for it. In any case, these headers should be added to the source, so that they provide consistent values across all languages. Alternatively, you can use the aforementioned Path Settings dialog to force the Proxy to override the cache headers and make the site cacheable, at the risk of diverging from the original, even if only for a short time.

Technical Reference

6.1 Architecture

6.1.1 Modularity

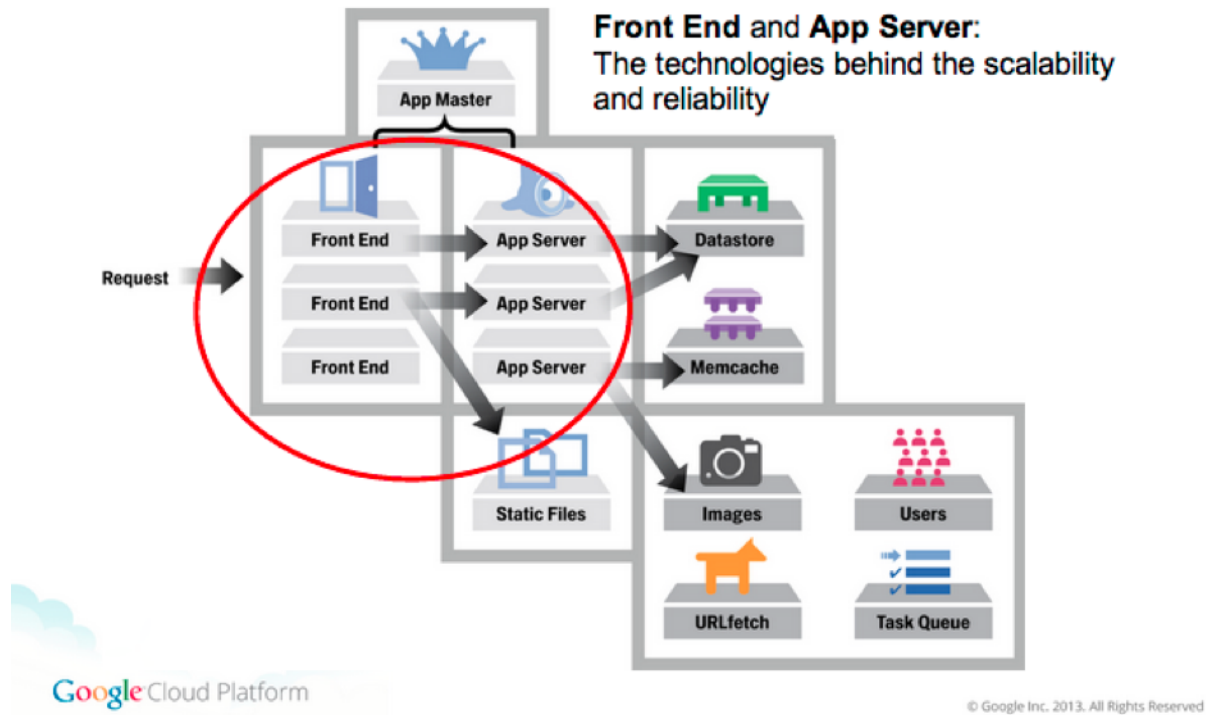
The Translation Proxy is based over Google's AppEngine infrastructure, split into frontend and backend modules. Each module encompasses variable numbers of instances, scaling automatically in response to demand. Modules are versioned and deployed separately, and can be switched independently, if needed.

Frontend instances serve requests from visitors to translated pages (in addition to serving the Dashboard and providing user-facing functionality). Requests are routed to the Proxy Application via the CNAME records created during the publishing process.

Backend modules are responsible for billing, statistics aggregation, and handling potentially long-running tasks, like XML import-export. Backend instances are not directly addressable, and provide no user-facing interaction.

6.1.2 Underlying technologies

Immediately underlying the Proxy Application is the AppEngine infrastructure, responsible for rapidly scaling the deployed application. AppEngine also handles communication with the Google Cloud Datastore, a high-replication NoSQL database acting as the main persistent storage; as well as the Google Cloud Storage system, an also-distributed long-term storage. Logging is provided by Google Cloud Logging, while BigQuery provides rapid search ability in the saved logs on request.



Encompassing the entire application is the Google EdgeCache network, proactively caching content in various data centers located regionally to the request originator. Any content bearing the appropriate headers (`Cache-control:public; max-age=/\d+/\` and `Pragma:public` - both are required) is cached by the EdgeCache for as long as needed, for requests originating in the same geographic area.

The current instance of the Proxy Application is hosted in the US central region, as a multi-tenant application (serving multiple users, but dedicated to proxying). However, single-tenant deployments (dedicated to a single user), special deployments to other regions (EU or Asia), or internal systems fulfilling the AppScale system requirements can be discussed on a per-request basis.

6.2 Request Handling

In the Translation Proxy, frontend instances are responsible for serving translated pages. Thanks to AppEngine's quick-reaction scaling, the number of frontend instances aggressively follows (and somewhat predicts) demand, keeping latency low. The general life cycle of a proxy request can be described as follows.

- The incoming requests, based on the domain name, reach the Google Cloud (rerouted via DNS record CNAME, pointing to `ghs.domainverify.net`).
- Based on the domain name and the deployed Proxy application, AppEngine decides that this specific request should be routed to the Proxy AppEngine deployment.
- The request reaches the Proxy Application internally; the application does a lookup against the domain for the associated project. There are special domain names, and the final serving domain, for which caching is activated.
- Based on the URL, the Proxy application determines the matching Page in the Proxy database. The database has a list of segments, pointing to our internal Translation Memory (TM). We retrieve all these existing Database entries, including the translations for the given target language.
- The Proxy application processes the incoming URL request, and transforms it to point back to the original site's domain. Then, the source content of the translation is sourced, according to cache settings in effect on the

project.

- If source caching is *disabled*, the application issues a request, and retrieves the result from the original web server, which is hosting the original website language.
- If source caching is *enabled*, a local copy (a previously stored version of the source HTML) is used, instead of issuing a request to the original web server.
- Depending on the `Content-type` of the response, the appropriate `Translator` is selected, and the response is passed to an instance of the `Translator` as a document. The behavior of the `Translator` can be affected by cache settings as well.
 - If binary caching is *disabled*, the application then builds the Document Object Model (DOM) tree of the result, finally iterates through all the block level elements, and matches them against the segments loaded from the database. If there's a match, we replace the text with the translation. If not, we 'report' it as a missing translation.
 - If binary caching is *enabled* and the hash of the source HTML *matches* the one stored in the cache, a previously prepared and stored translated HTML is served.
 - If binary caching and keep cache are *both enabled*, and the hash of the source HTML *doesn't match* the one stored in the cache, the proxy translates the page using the TM. If the number of translated segments is higher than the previously prepared and stored translated HTML, the new version is served; otherwise the old one. (Keep cache can be thought of as a "poor man's staging server").
- Hyperlinks in the translated content are altered by the `LinkMapper` to point to the proxied domain instead of the original. This affects all `href` or `src` attributes in the document equally, unless the element is given the `__ptNoRemap` class. At this point, resources may be replaced by their localized counterparts on a string-replacement basis.
- The application serializes the translated DOM tree, and writes it to the response object.
- Before final transmission takes place, the Proxy may rewrite or add any additional headers, such as `Cache-control` or `Pragma`.
- Finally, the Proxy serves the document as an HTML5 stream, as a response to the original request. AppEngine *must* close the connection once the response is transmitted, so proxying streaming services is not possible in this fashion!

6.3 Classification of Content

The Translation Proxy distinguishes two main types of content: text content and resources. The key difference is that text content may be translated, while resource content is treated opaquely, and only references can be replaced as resource localization. It is possible to reclassify entities from Resource to Text, but not the other way around.

During proxying, resources are enumerated, and any already-localized references are replaced, while text content is passed to an applicable `Translator` implementation for segmented translation.

6.3.1 Text content

By default, the Proxy only handles responses with `Content-Type:text/html` as translatable. To process HTML content, the source response's content is parsed into a `Document`, then text content is extracted from the DOM-nodes. Additionally, various attribute values are processed (without additional configuration, `title` and `alt`).

The content is then transformed into `SourceEntry` entities server-side. Each block element comprises one source entry, with a globally unique key. If segmentation is enabled on the project, the appropriate rules are loaded (either using the default segmentation or by loading the custom SRX file attached to the project), and the content is segmented

accordingly, with the resulting token bounds stored in the `SourceEntry`. Along with the `SourceEntry` entities, the corresponding `TargetEntry` and `SourceEntryTarget` entities are created. `TargetEntry` entities, as the name suggests, hold the translations; `SourceEntryTargets` act as the bridge between the two, and hold the segment status indicators for both.

The content of source entries is analyzed in the context of the project, and statistics are computed. These statistics include the amount of repeated content at different confidence levels based on the similarity of the segment - The Translation Proxy differentiates five levels of similarity:

1. 102%: Strong contextual matches: every segment in the block level element (~paragraph) is a 101% match, where all the tags are identical. These matches do not result in the creation of new `SourceEntry` entities, thus changes in one place are propagated instantly to all occurrences.
2. 101%: Contextual matches: both tags in the segment, and contexts (segments immediately before and after) match.
3. 100%: Regular matches: the segment is repeated exactly, including all tags.
4. 99%: Strong fuzzy matches: tags from the ends are stripped out, words lowercased, numbers ignored.
5. 98%: Weak fuzzy matches: all tags are stripped out (may have to be adjusted manually afterwards!), words lowercased, numbers ignored. If the Proxy cannot match the tags between the translation and the source due to excessive differences, all tags are placed at the end of the segment, requiring manual review!

These classifications are reused during memory-powered pre-translation in order to select the best applicable translation or to propagate existing translations.

6.3.2 Resource content

By default, any content with content types other than `text/html` are treated as a resource, and is not a candidate for translation, only replacement en bloc. This mainly includes `application/javascript`, `text/xml`, and various `image/*` content types. Every resource can be given different replacements per target language, and if required, certain resources (`application/javascript` and `text/xml`) can be made translatable after pre-configuration is done. In this case, instead of references being replaced, the appropriate `Translator` will be instantiated and the content passed to it. This can enable partial or complete translation of dynamic content transmitted as JSON or XML.

6.4 Translation Memories

The Translation Proxy can be configured to maintain and leverage internal translation memories. These memories can contain more than one target locale allowing leveraging them for any pair of locales contained within.

As opposed to project dictionaries, translation memories are keyed to the *user* creating them, and can be assigned to any project the user has Backup Owner privileges or higher. Any project can contain an arbitrary number of memories, but one must always be designated the default: only this memory will be utilized when segments are being written; while pre-translation and suggestions are fed from *all* memories assigned to the project with applicable locale configurations.

6.5 Using TMs

Translation memories are initialized empty, and must be first configured with locales. After the target locales are defined, the memory can be populated. There are three ways a segment can be injected into the memory:

- TMX-import: The Proxy can digest a standard TMX (Translation Memory eXchange) file and populate a designated memory based on its contents. The memory must be configured with at least one of the target locales of the TMX file. Duplicate segments are either merged (if for different locales) or discarded during import.

- **Project population:** The Proxy can populate the memory from the project it is currently assigned to. The memory must be configured with at least one of the *project's* target locales for this to work. If there are several locales assigned to the memory, the UI will treat them as a set, and offer the *intersection* of the memory and the project's locales as the default. This set can be further restricted by removing locales from the population task before committing it. This action is logged in the project's Audit Log.
- **Individual Injection:** If a memory is assigned to the project with at least one locale present on both, it will be available on the Workbench for use. Confirming one or more segments will trigger the `saveToMemory` action, injecting the segment in its current form into the memory.

Memories are used for two tasks on the UI:

- Pre-translation tasks can leverage any memories assigned to the project, provided the memory is configured with the correct locale. This applies to user-triggered Pre-translation, as well as Automatic Pre-translation triggered by new content. Only content with confidence levels above the user-configured threshold will be used, matches with lower percentages are discarded.
- The Workbench automatically leverages any memories with the appropriate locales on segment selection. Matches are displayed in the Suggestions tab of the sidebar, along with their match percentages. Additionally, all memories on the project with the applicable target segments can be queried at will by entering a search term.

6.5.1 Confidence levels

The Proxy differentiates five levels of similarity between individual segments/entries (see here). Memory application yields the best results between 101% and 99% - 98% matches disregard tagging, and may need manual adjustment. However, searching below 98% is also possible, using the Google Search API, but these matches should be used with caution, as there is no guarantee regarding their accuracy due to the Search API's word stemming.

6.6 Page modifiers

Due to the way the Proxy Application operates, it becomes fairly easy to modify the pages as they are being served. Because the datastream must pass through the proxy to have the translation embedded, the Proxy Application can insert JavaScript modifiers, modify style sheets, and even embed entire pages that do not exist on the original.

- **CSS Editor:** the Proxy Application can be used to insert locale-specific CSS rules into the site being served. The rules are inserted as the last element of the `head` on every page served through the proxy. The most common use of this feature is to alter the writing direction for non-Latin scripts, such as Arabic.
- **JavaScript Editor:** the JavaScript edited here is inserted into the `head` element of every page being served through the Proxy Application. As the last element of the `head`, it has access to any global variables added by scripts before it.
- **Content Override:** the Proxy Application can create a “virtual” page in the site or override an existing one with custom code. For any requests to an overridden page, the corresponding remote server request is not sent, and the override contents are used as the basis of the translation. The source is not required to be HTML, custom content-types can be entered, along with customized cache headers, and status codes (HTTP status codes are restricted to those permitted by the [Java Servlet class](#)!) - note that the 300-family of status codes requires the `Location` header to be defined as well.

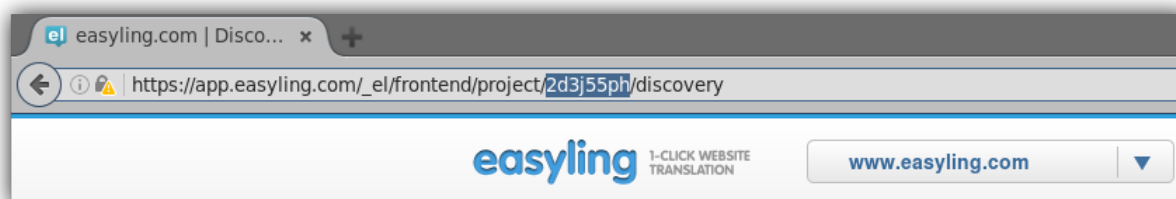
Both the CSS and JavaScript injectors can use already-existing files for injection instead of copied content. The injected files must be handled by the project in some way (either by being in the project domain, or in the domain of a linked project), or be created by a content override. The order of definition for these entries matters, as they will be inserted into the document in the order they are displayed on the UI, which may cause dependency or concurrency issues!

Troubleshooting & Support

7.1 Contacting Support

Website translation can be a complex task, even with the help of a piece of software like the proxy. Finding the root of the problem can be equally complex for those of us on support duty. Therefore, we would like to give you a few pointers on what information to supply if you decide to contact us.

This first thing we need is the **project code**. This eight-character unique string uniquely identifies the project. As you can see in the screenshot below, the project code is located in the address bar of your browser - you can copy the entire URL for us, but the project code by itself should suffice.



We also need a **thorough description of the problem**. Screenshots are tremendously helpful, especially if you have layout issues on the translated site. If you're running into translation issues, please give an example segment along with the page link it can be found on.

If you have issues with importing XLIFF or TMX files, please attach them so that we can take a look. If you have questions about statistics, reports, or crawl logs, attaching or linking them in your query will considerably speed things up for us.

The information you provide will help us uncover the root cause of the problem. This often requires a bit of “detective work” on the original site, so we ask for of your patience while we figure things out. Someone from the support team will respond shortly with a solution, a request for more information, or simply an update.

7.2 Issues

7.2.1 Scanning Content Behind Secure Login

To scan content behind secure login, please follow this procedure:

1. Open your project and navigate to the **Content** menu.
2. Open the **Pages** list.

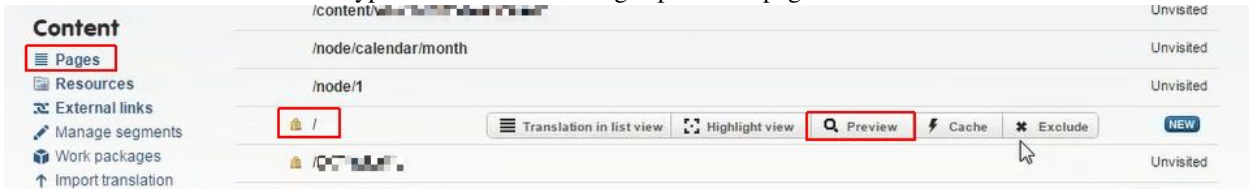
3. Visit the page with the login, if it is listed, and click **Preview**.

OR

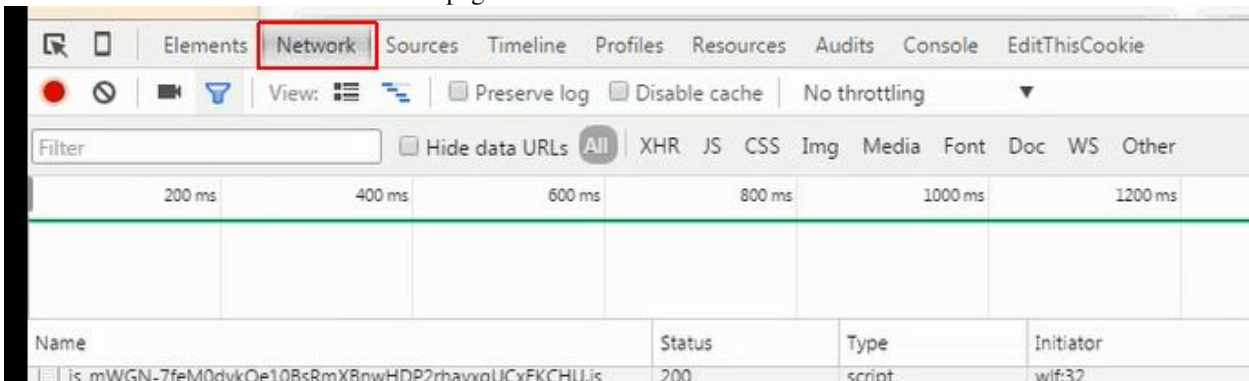
Go to the **Preview** of the front page (the “/”, the first one on the Pages list). It will give you the front page through the proxy.

OR

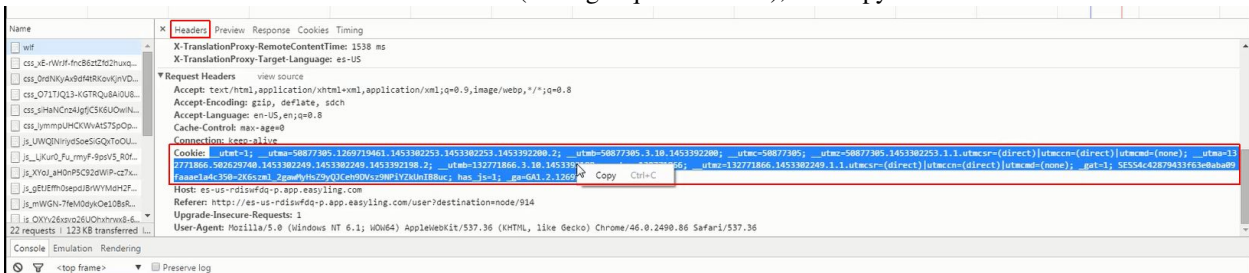
Go to the address bar and type in the URL of the login-protected page.



1. Enter your login details.
2. Open your browser's DevTools from the Menu (F12 on Windows).
3. Go to** Network** and reload the page.



1. Scroll up to the first item and click on it.
2. Under headers scroll to the cookie header (among request headers), and copy the entire header.



1. Pass it to the Proxy: go back to your project and click on **Content**. Paste the entire content in the **Scanning cookies** field.

selected pages, retrieve the content for translation and store it in the database. Subsequent scanning will also discover new pages and new/changed content on the site.

Settings

☐ Process pages in source language only (requires Google language detection API)

Ignore queries

Group pages by ignoring query parameters
Matching source segments are merged across pages in the group

Scanning cookies
These values are not saved for security reasons. [More about cookies](#)

Scan page limit number of pages maximum Leave blank for Unlimited

Look for changes or [Scan manually now](#)

Last scan: Started at 0:13 Jan 20, 2018

1. Click on **Scan manually** and specify the required scanning settings. You will receive an e-mail notification once scanning is ready and new content is available for translation.

7.2.2 XLIFF import error

When you import your translated XLIFF file back to easyling, you will receive an e-mail notification when the process is ready. This mail contains the URL of the import log, and an overview of the log entries:

- Error:
- Warning:
- Info:

If you see other than '**Error: 0**' in your notification mail, the XLIFF file needs fixing. Usually these are tag placement errors that can be easily fixed in a text editor like Notepad++ or Sublime (the ones that have syntax highlighting, to make this fixing job easier), yet they do need attention, as the corresponding translation will not show up on the website.

- Open the XLIFF in your text editor
- Open the log file and check the error message(s)
- Do the necessary corrections in your XLIFF (see 'Troubleshooting')
- Save & upload the corrected XLIFF

In very serious cases the import might fail completely, but this is very rare. These cases include: attempt to upload an XLIFF related to another project, XLIFF with target language that doesn't exist in the project, and fully invalid XML in the file. In most cases the file is imported, only the faulty entries are omitted.

Please note that you need XLIFF files. Ideally, the export format of the CAT-tool should be the same as the import format, and as you import an XLIFF file for translation, the output should also be a standard XLIFF file. However, some versions of Studio tend to create an **SDLXLIFF** file upon exporting the translation. In this case, simply use the "**Finalize**" batch task or open the document in the **Editor**, press **SHIFT+F12** and select the target file location. This will create the XLIFF file for you (instead of SDLXLIFF).

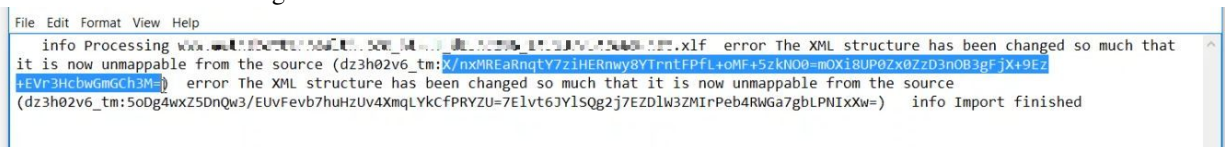
You might also need to disable segment info storage in Studio (*Options -> File Types -> XLIFF -> Settings -> 'Do not store segmentation information in the translated file'* should be checked). This may require creating a new project.

Troubleshooting

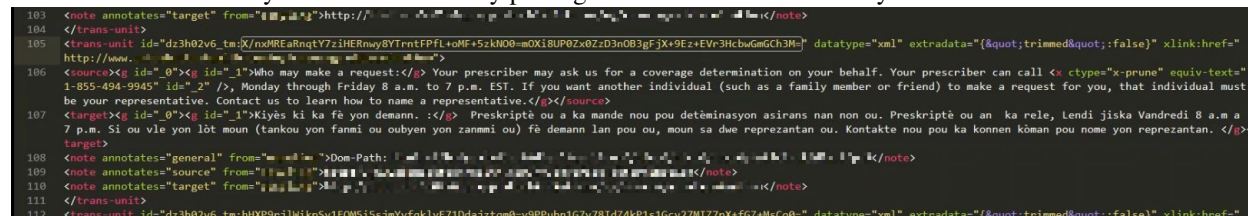
Error: *The xml structure has been changed so much that it is now unmappable from the source*

Fix:

1. Open both the **XLIFF** file and the **error log** in a text editor.
2. Select & copy the **TM-key** of the faulty entry, the part after ‘(trans-unit id=”xxxxx”)_tm:’ in parenthesis right after the error message



1. Search for this key in the **XLIFF** file by pasting it into the ‘Find’ field. Only 1 translation unit will match.



1. Compare the tags in the source and target languages, and fix the mismatch by editing the **target text**. (You can also use an online text comparison tool for this task: copy-paste <source> ... </source> in one pane and <target> ... </target> in the other one.)
2. Save the corrected XLIFF file and upload again. It should give no error message now.

OR, alternatively,

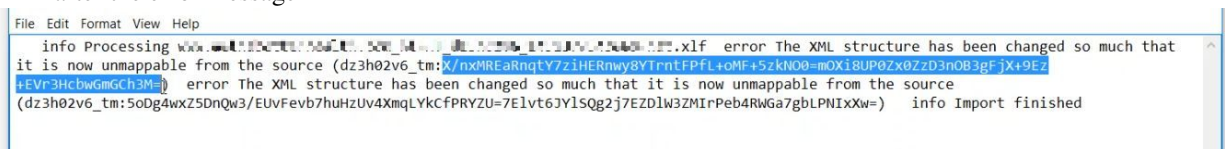
1. Go back to your CAT tool, where you did the translation and open the faulty file for editing
2. Run QA. It will list you all the tag mismatches
3. Navigate to the faulty segment(s) and fix the tags
4. Export the corrected file and upload it. It should give no error message now.

Error: Content found outside of outermost element

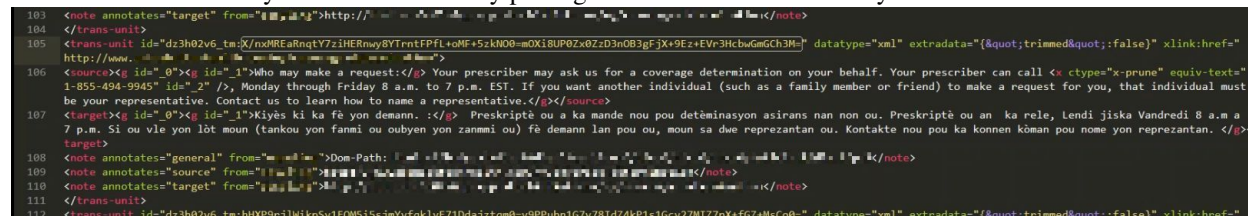
Fix:

Practically this means that there is an extra space before the starting <target><g*> or after the closing </g></target>.

1. Open both the **XLIFF** file and the **error log** in a text editor
2. Select & copy the **TM-key** of the faulty entry, the part after ‘(trans-unit id=”xxxxx”)_tm:’ in parenthesis right after the error message



1. Search for this key in the **XLIFF** file by pasting it into the ‘Find’ field. Only 1 translation unit will match.



1. Delete the extra space around the tags
2. Save the corrected XLIFF file and upload again. It should give no error message now.

OR, alternatively,

1. Go back to your CAT tool, where you did the translation and open the faulty file for editing
2. Run QA. It will list you all the tag mismatches
3. Navigate to the faulty segment(s) and fix the tags
4. Export the corrected file and upload it. It should give no error message now.

IMPORTANT! Most of these issues can be avoided if the QA parameters of your CAT tool are set up properly, and you run QA before exporting your XLIFF files. Please make sure to check your translation for tag consistency and extra spaces; these are critical errors in website translation that can spoil the code.**

Error: *Illegal character*

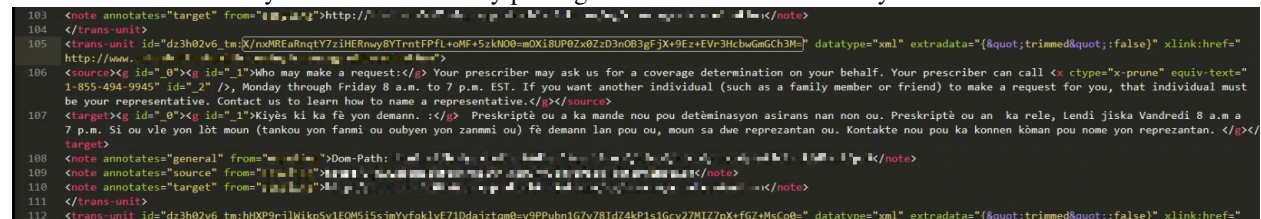
Fix:

The reason for this error is usually a coding mismatch: all our XLIFF files are exported using the world-standard UTF-8 codepage. However, your CAT tool may save the file using another codepage, depending on the language, which may cause certain characters to appear as invalid.

1. Open both the **XLIFF file** and the **error log** in a text editor
2. Select & copy the **TM-key** of the faulty entry, the part after ‘(trans-unit id=”xxxxx)_tm:’ in parenthesis right after the error message



1. Search for this key in the **XLIFF file** by pasting it into the ‘Find’ field. Only 1 translation unit will match.



1. Check if you see any strange characters, like squares or other meaningless characters.
2. Go back to your CAT tool and change the export options to use UTF-8 encoding. (As UTF-8 is a universal standard, it should be available.)
3. Re-export the XLIFF with UTF-8 encoding and upload it. It should give no error message now.

7.2.3 Publishing issues

Translated page doesn't show up

Issue: *I've just uploaded the translation of some new pages, but they don't show up on the translated site / they still appear in the source language.*

Fix:

You experience this problem because Target cache is enabled, and you need to clear the Cache to update content to be served. The very reason for using a Target cache is to mask unfinished translation and avoid bleed-through. The Target cache shows the last fully translated version of the pages - so if content changes on the original page, it remains hidden on the translated site until a fully translated version is available.

To fix the issue you need to explicitly delete the page from the Cache, so that the updated content could be loaded in upon the very first viewing of the page.

If you have several languages, it might be more convenient to clear the entire Target cache by clicking the Trash icon.

Translated page is not listed in Google search results

Issue: *The pages served from the Proxy are not listed in Google search results.*

Fix:

Google has never indexed the site in the first place, most likely due to the fact that there are no “hreflang” links on any of the pages, so the Googlebot has no idea there are other pages to look for. More information on the element and how it affects Google rankings may be found at <https://support.google.com/webmasters/answer/189077?hl=en>

Additionally, creating and submitting a sitemap (more information at <https://support.google.com/webmasters/answer/2620865>) to Google in order to force an indexing of the pages can also help. Even so, without the hreflang attributes, it may mean that some penalty is applied to the rankings, due to perceived duplicate content.

7.2.4 Using the hreflang Element

Google’s CrawlerBot will eventually find your translated page if there are any links to it. However, if the content there is not marked appropriately, it will not be given the same SEO scores as your main content. In fact, it may even be treated as duplicate content, and a scoring penalty may be applied.

To prevent this from happening, you need to provide the GoogleBot with information on how the translated sites relate to the original. The easiest way to do this is the `<link rel="alternate" hreflang="" href="">` element.

These elements have to be placed in the page head (i.e. before the HTML body), and have two rules that must be satisfied in order for the GoogleBot to consider them:

- hreflang elements must be reciprocal: if a link points to a translated site, the translated site must point back to the original as well.
- hreflang elements must be circular: each language must also refer to itself with a link.

Consider the following HTML snippet from an imaginary site at `http://example.com`:

```
<html>
<head>
<title>Title Here</title>
<link rel="alternate" hreflang="en" href="http://example.com" />
<link rel="alternate" hreflang="jp" href="http://jp.example.com" />
</head>
[...]
```

and its translated counterpart at `http://jp.example.com`:

```
<html lang="ja-JP">
<head>
<title>Title Here</title>
<link rel="alternate" hreflang="en" href="http://example.com" />
<link rel="alternate" hreflang="jp" href="http://jp.example.com" />
```

```
</head>
[...]
```

This snippet will provide proper SEO, since it satisfies both criteria: the references the English and the Japanese site are reciprocal (one refers to the other and vice versa) and the references are also circular (both languages also refer to themselves as well as their counterparts). This provides all the information GoogleBot needs to index each site in its rightful place and apply SEO scores across both domains.

For more information on this topic, see this article from Google: <https://support.google.com/webmasters/answer/189077?hl=en>

7.2.5 WPEngine issues

Due to WPEngine's *caching system* and *Redirect bots* settings you might experience any of the following issues on your **WPEngine hosted sites**:

1. Scan extracts outdated content from WPEngine cache
2. Scan returns **301: Moved permanently** error message for existing pages
3. Translated page is redirected from HTTPS to HTTP - which results in an error due to mixed content

WPEngine caching uses different so-called *buckets* based on request type, and they have one for bots. If the request comes from Google or other listed user-agents, and/or the URL has `?ver=` followed by a number, the redirect bots settings take effect.

The above issues can be resolved on the WPEngine site by turning off the *redirect bots*.

7.2.6 Captchas

Captcha doesn't work on the translated site

This issue is most likely results from page caching. Certain captcha solutions, like WP plugins are hardcoding the image URL into the HTML, instead of sending it asynchronously. During the crawl that builds up the Source cache for the project one hash is saved, so it becomes static, while it should be changing on each occasion. As a result, the server rejects the request because of the outdated verification image.

Fix:

- disable caching altogether to make captcha work
- use caching without captcha
- use another CAPTCHA system that uses async requests to retrieve the verification image, like Google's Re-Captcha solution

Another possible cause can be the CORS-header. If the proxied page is not listed as an allowed origin, the browser blocks the page when it tries to load the image.

8.1 General

8.1.1 Some parts of a site are on a subdomain. How will the crawler pick them up?

The sites `www.company.com` and `blog.company.com` are treated as separate domains by the crawler. From the vantage point of a crawler running on `www.company.com`, a path on `blog.company.com` is an external and will be treated as such. The solution is to create two separate projects and link those with each other.

8.1.2 The Discovery went beyond the limit I set. Why?

A crawl will finish the current round and visit the redirects and links on the last page. If it took the limit too literally, that could potentially result in trailing links being thrown out.

8.1.3 Can I get page-specific repetition statistics?

Repetition statistics make the most sense in a site-wide context. The problem with controlling calculations on a per-page basis is that it is not true to life to call a segment on a given page a “canonical instance”. Take a navigation bar or a footer, for example. It will be “repeated” on all pages, but it cannot be said to “belong” to any one of them. Easyling stores the first instance it comes across and then propagates its translation to all other instances.

8.1.4 The page I’m trying to translate has prices. What can I do to handle local currencies?

The prices themselves can be made translation-invariable, but real-time price handling for different currencies will have to be implemented by the client on the source site, making it possible for the proxied site to access the locale-specific information. Pricing of products and services also has legal / market implications that are beyond the tasks of LSPs. Of course, once currency-specific information is accessible from the original site, we are happy to help with integrating any backend calls / ajax requests on the proxy.

8.1.5 How do I enable automated scan on my project?

To enable automated content extraction on your site, please go to Content, and choose either of the daily, weekly or monthly options in the drop-down next to the Look for changes option.

8.1.6 Is it possible to set up automated scanning behind secure login?

No, scanning can't be automated behind secure login. For such processes you need to extract cookies with you browser's dev. tools and pass them on to the proxy. Some cookies get invalidated over time, and we don't store cookies either.

8.2 Caches

8.2.1 Can I preview newer content on the workbench without causing bleedthrough on the published site?

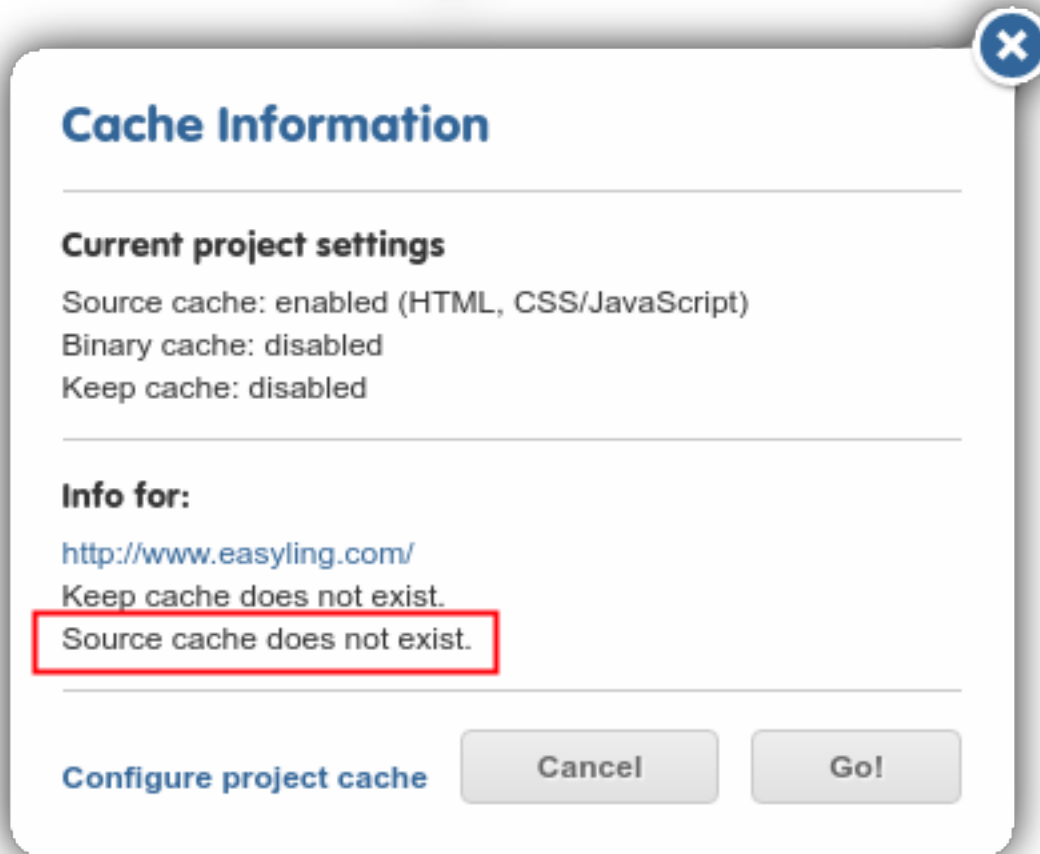
You can customize which Source Cache to use on which proxy mode - go into Page Cache, choose custom settings and select Disabled from the dropdown menu. The preview mode will display all new content. It is recommended that you keep TM Freeze turned on while exploring the new content, otherwise everything will be automatically added to the Workbench.

8.2.2 Does building a Source Cache cost any money?

You can use Content Scan with the appropriate options checked to build your Source Cache. As long as there is no new content to pick up, Scan costs the same as a Discovery.

8.2.3 How can I check if a page uses the Source Cache?

Go into **Pages view** in the Dashboard. If you hover with the mouse over a page in the list, you will see a **Cache** button. Click on it to verify Source Cache information for that page. If there is no Source Cache for that page, you will see the following screen:



8.2.4 Does building a Target Cache cost any money?

Setting aside the inherent cost of the Page Visits you have to accrue to build them, Target Caches are free of charge.

Glossary

We use a set of recurring terminology in this manual and also in our Support Channels - we collect them here for your reference.