

NEC TM UI Description

Version 1.00



This report is the result of the Connecting Europe Facility (CEF) programme

Action No: 2017-EU-IA-0149

Agreement number: INEA/CEF/ICT/A2017/1565776

0. Revision Control	3
1. INTRODUCTION	3
1.1 Purpose	3
1.2 Scope	3
1.3 Overview	3
1.4 Definitions and Acronyms	4
2. UI OVERVIEW	4
2.1 Website UI	5
2.1.0 Login	5
2.1.1 Fuzzy Match Services	5
2.1.1.1 Search translation memory segments	5
2.1.4 Import Services	7
2.1.4.1 Import translation memory segments from TMX file	8
2.1.5 Export Services	8
2.1.5.1 Export translation memory segments to zipped TMX file(s)	9
2.2.5.2 Download exported file or list all available downloads	9
2.1.6 Stats Services	11
2.2 REST API	11
2.2.1 Fuzzy Match Services	11
2.2.1.1 Search translation memory segments	11
2.2.1.2 Batch search translation memory segments	14
2.2.2.6 Authorization	17
2.2.3.3 Get tag details	18
2.2.4 Import Services	19
2.2.4.1 Import translation memory segments from TMX file	19
2.2.4.2 Add new translation memory unit	20
2.2.5 Export Services	21
2.2.5.1 Export translation memory segments to zipped TMX file(s)	21
2.2.5.2 Download exported file or list all available downloads	24

0. Revision Control

Author	Description
Alex Helle	Initial Version of Specification (April 24th 2019)
Amando Estela	Initial Revision (April 25th)

1. INTRODUCTION

1.1 Purpose

This user UI document describes the web and API interface of NEC TM based on ActivaTM, cloud-based Translation Memory tool. The target audience are users of the tool.

1.2 Scope

ActivaTM is a fast, highly-scalable cloud-based Translation Memory tool developed by Pangeanic as a part of the EXPERT EU project. The goal of ActivaTM is to seamlessly integrate all available TMs to serve vendor-independent tools in the standard translation flow. ActivaTM integrates (via plugins) with multiple manual and machine translation software (SDL Studio, PangeaMT, NeuralMT, Bing Translate).

NEC TM Data project will provide the centralised infrastructure for efficient data sharing, TM matching, TM retrieval, and tag categorisation of resources generated by Member States/ EEA. This will enable the development of NEC TM, an open source software developed based on Pangeanic's translation memory database ActivaTM.

1.3 Overview

The document describes the web and API interface of NECTM.

1.4 Definitions and Acronyms

TM - Translation Memory

ES - ElasticSearch

CAT - Computer-Assisted Translation

Translation unit - data structure, containing two (or more) text segments of different language along with additional properties. Sometimes used interchangeably with segment

Segment - single-language text string, part of translation unit

API - Access Point Interface

UI - User Interface

2. UI OVERVIEW

A Translation Memory (TM) is a collection of segments helping to improve quality of translation process in CAT tools (such as Trados Studio) by reusing translation of phrases, sentences or even paragraphs (commonly known as segments) from previously known translations. Usually, automatic machine translation tools use translation memory in-memory (loaded from standard TMX format) or by using web service. The number of segments can be huge if all possible European language combination pairs are taken into account. Current non-commercial solutions appear to be either non-scalable (such as dealing with TMX files) or in early stages of development, such as TinyTM database. This document describes the user UI of NECTM, a fast and scalable cloud-based TM database with a goal to overcome the limitations of current solutions for TM storage & query.

There are two ways to access NECTM, using the Website UI and using the API REST.

Website UI is the web service that facilitates the interaction with NECTM content (adding, deleting, updating, cleaning translation units etc.), managing users and managing background jobs through a web browser.

REST API (*based on Flask¹*) is a programming interface for interacting with NECTM content (adding, deleting, updating, cleaning translation units etc.), managing users and managing background jobs. The user is authorized by accessing authorization endpoint (with username and password). Access to other endpoints is granted based on JWT token return by authorization endpoint and depending on user role (admin or user).

¹ "Flask." <http://flask.pocoo.org/>. Accessed 23 April 2019.

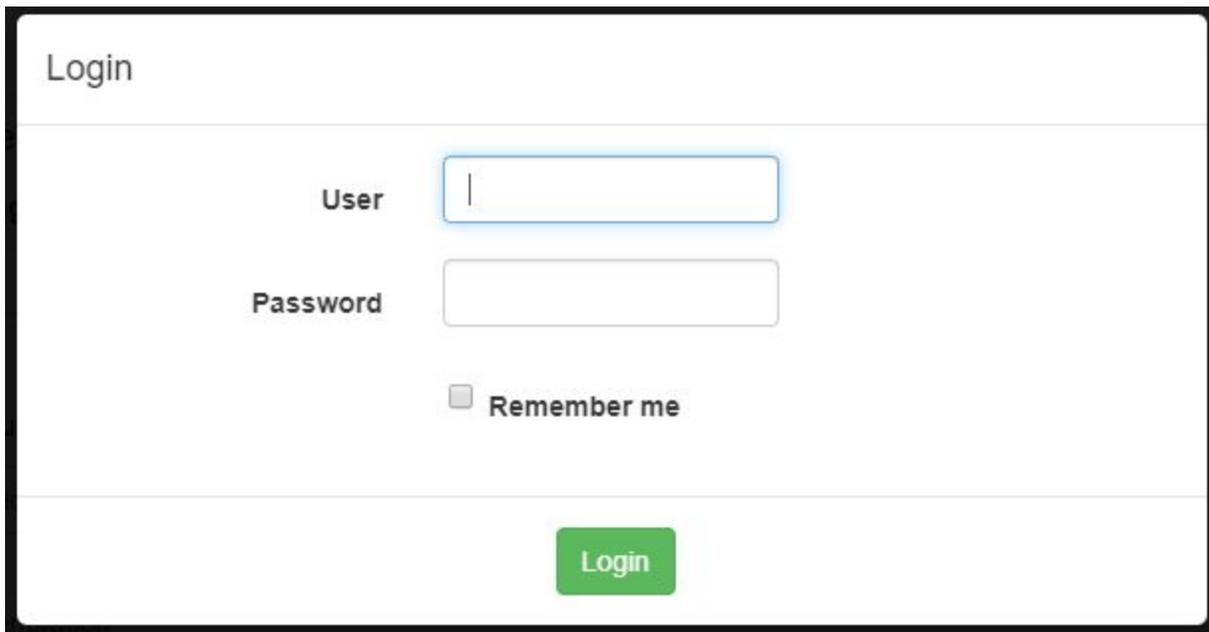
2.1 Website UI

This section describes the main Website UI functions for the services in NECTM:

2.1.0 Login

The default URL access is `http://ip-installation:27979/admin/` (27979 is the default port used in the installation).

After accessing Website user UI a login will be shown:



A valid User/Password shall be provided. If the “Remember me” option is checked a “login cookie” is issued in addition to the standard session management cookie.

2.1.1 Fuzzy Match Services

Services related to query for translation units stored in NECTM.

2.1.1.1 Search translation memory segments

Permission: user

Go to the menu “Translation Memory” and the tab “Query”.

Type the text to query in the text field, select the pair of languages and click on the button “Query”.

Match	Source	Target	Tags	Username	File names	Update date
10	The European Central Bank	The European Central Bank	ECB	admin	['ECB_en-fr.tmx']	12/04/2019 12:04:13
10	Avis sur la modification du statut du Bank Ċentrali ta » Malta / Central Bank of Malta en vue de l'adoption de l' euro (Opinion on amendments to the Central Bank of Malta 's statutes in preparation for the adoption of the euro (CON /	ECB	admin	['ECB_en-fr.tmx']	12/04/2019 12:04:15

The language pairs that appear are the ones used by the TMXs already uploaded, so we'll get a new specific language pair uploading a TMX that contain the specific language pair.

You'll get results depending on the options and filters selected. Please refer to the parameters on section [2.2.1](#) for more information about the options/filters that can be selected.

The main filter is the usage of the filter "TAG" (refer to section [2.1.3](#)), filter used during the import (refer to section [2.1.4](#)) to classify the translation units. The users only has access to the tags they have defined in their scopes.

Translation Memory

Query Import Export Generate Delete Maintain Tags Stats

Query TM

bank

Source language Target language

FR EN

Options Filter by... Query

TAG None selected

Results

It will show a "Forbidden error" if we try a tag combination that isn't allowed in the scopes of the user (refer to section [2.1.2.3](#)).

2.1.4 Import Services

Services related to the import of content in NECTM. There is only one way through the website: TMX.

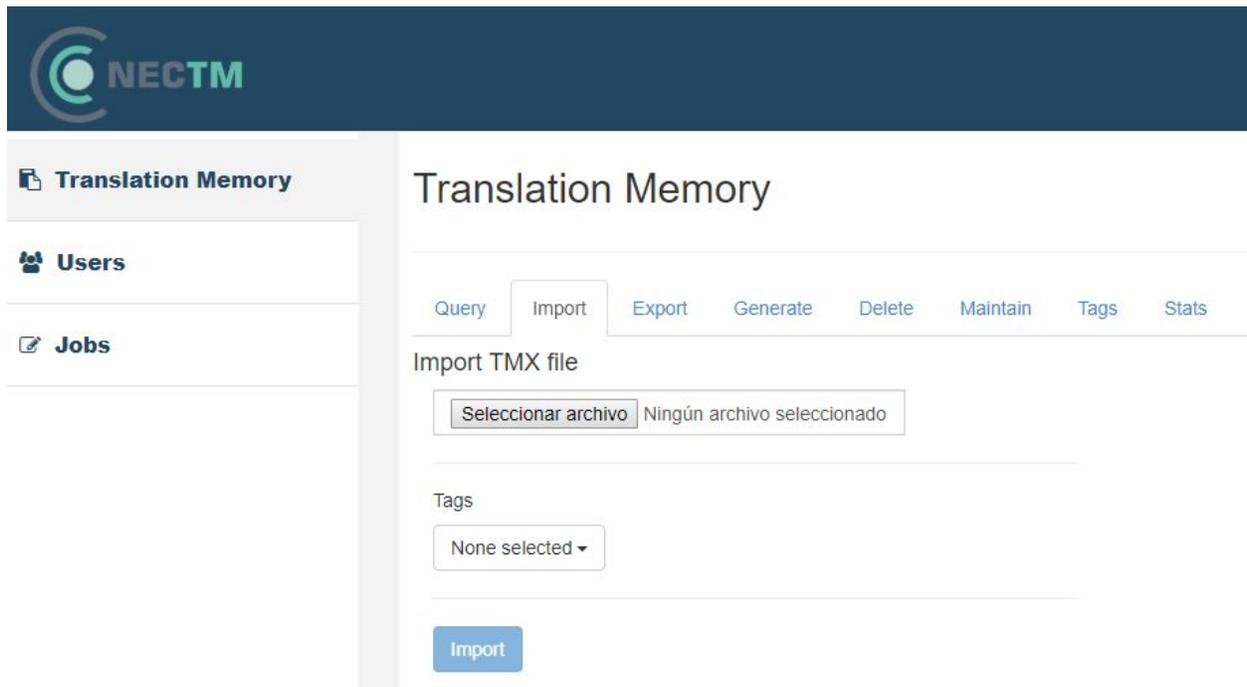
It's possible to specify tags for categorization of the content (refer to section [2.1.3](#)).

We can import a single translation memory unit through the REST API (refer to section [2.2.4.2](#)).

2.1.4.1 Import translation memory segments from TMX file

Permission: admin, users authorized

Go to the menu “Translation Memory” and the tab “Import”.



Select a TMX 1.4b file format (or you can upload several if they are compressed in a zip file) to upload, select the “tags” to classify the content and click on the button “Import”.

Content that has been characterized with “private” tags can only be accessed by users with these tags specified in one “scope” of the user. Content that is flagged with “public” tags can be accessed by all users. The tags for “unspecified” type (example: legal, automotion, health...) can be selected by all Users, but the content will only be seen if it has a “public” type domain or “private” type domain if the User has access in the Scopes (refer to section [2.1.2.3](#)).

It will show a “Forbidden error” if we try a tag combination that isn’t allowed in the scopes of the user (refer to section [2.1.2.3](#)).

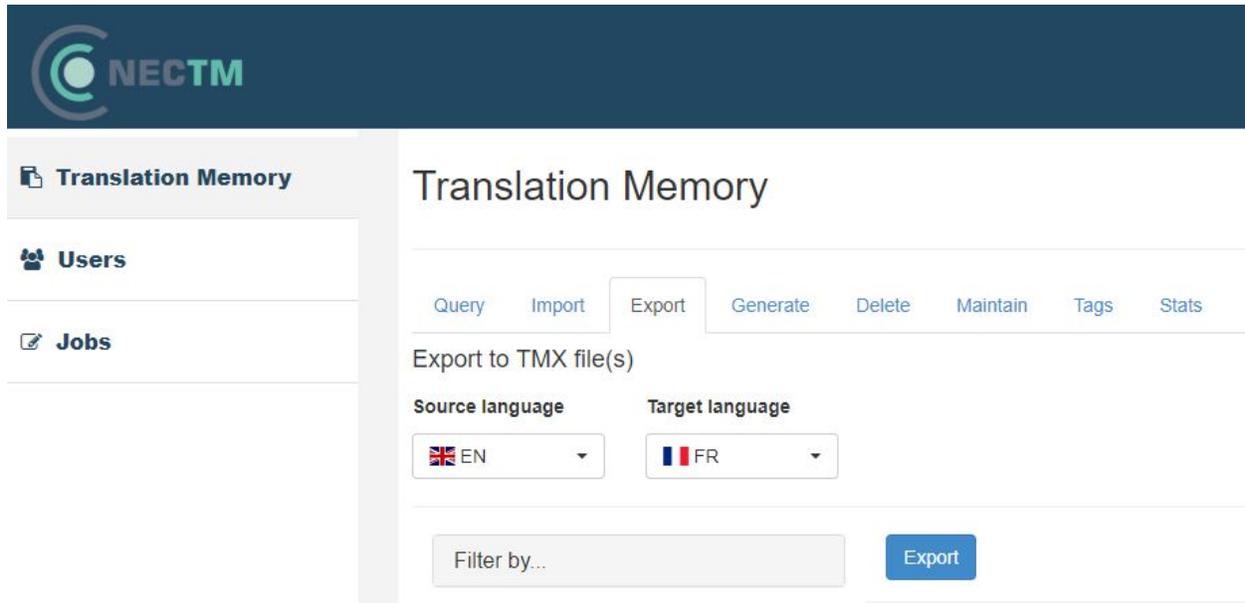
2.1.5 Export Services

Services related to the export of content in NECTM to a TMX file.

2.1.5.1 Export translation memory segments to zipped TMX file(s)

Permission: admin, users authorized

Go to the menu “Translation Memory” and the tab “Export”.



Select the desired filters and click on Export button. Please refer to the parameters on section [2.2.5.1](#) for more information about the filters to select.

The export will run on background and will generate a single zip file that it would be possible to download later. The export process can take minutes or hours depending on the amount of data to be exported and the hardware of the server. The export processes can be seen and cancelled in the Jobs menu (refer to section [2.1.7](#)).

It will show a “Forbidden error” if we try a tag combination that isn’t allowed in the scopes of the user (refer to section [2.1.2.3](#)).

2.2.5.2 Download exported file or list all available downloads

Permission: admin, users authorized

Go to the menu “Translation Memory” and the tab “Export”.

The screenshot shows the NECTM Translation Memory interface. On the left is a sidebar with navigation options: Translation Memory, Users, and Jobs. The main area is titled "Translation Memory" and has several tabs: Query, Import, Export (selected), Generate, Delete, Maintain, Tags, and Stats. Below the tabs, there is a section for "Export to TMX file(s)" with dropdown menus for "Source language" (EN) and "Target language" (FR). At the bottom of this section is a "Filter by..." input field and an "Export" button.

In the section “Manage Exports”, all the exports that have finished will be shown with a link to download.

This screenshot shows the "Manage Exports" section of the NECTM Translation Memory interface. It features the same navigation tabs as the previous screenshot, with "Export" selected. Below the tabs, there are language selection dropdowns for "Source language" (EN) and "Target language" (FR), and a "Filter by..." input field with an "Export" button. The main content is a table with the following data:

	ID	FILENAME	SIZE	EXPORT_TIME
	9d103adb-52d1-4b02-9184-af5a070810ad	EN_FR.zip	0.1 M	24/04/2019 11:23:32

Below the table, it says "Showing 1 to 1 of 1 rows".

It is recommended to remove the download once has been downloaded.

2.1.6 Stats Services

Permission: user

Go to the menu “Translation Memory” and the tab “Stats”.

Source	Target	# TUs	Tags
FR	EN	1273	ECB 1273
EN	FR	1273	ECB 1273

Showing 1 to 2 of 2 rows

It is shown different statistics related to the data:

- “Languages”: amount of TUs per language pair
- “Tags”: amount of TUs per tag
- “File Names”: amount of TUs per filename used during the import of a TMX file

2.2 REST API

This section describes the main REST API functions for the services in NECTM:

2.2.1 Fuzzy Match Services

Services related to query for translation units stored in NECTM.

2.2.1.1 Search translation memory segments

GET -> <https://localhost/api/v1/tm>

Permission: user

Field	Type	Description
token	String	Token returned by auth endpoint.

Parameter

Field	Type	Description
q	String	String to query
slang	String	Source language
tlang	String	Target language
smeta optional	String	Source metadata (stringified valid JSON). For example: <i>JSON.stringify({"x-context-post": "pretag", "x-context-pre": "posttag"})</i>
tmeta optional	String	Target metadata (stringified valid JSON)
out optional	String	Output format. Default value: json Allowed values: 'json', 'moses'
limit optional	Number	Limit output to this number of segments

min_match optional	Number	Return only match above or equal to given threshold (0-100) Default value: 75
strip_tags optional	Boolean	Strip all XML tags from the query Default value: false
concordance optional	Boolean	Concordance search mode Default value: false
aut_trans optional	Boolean	Apply machine translation if match score is less than a threshold Default value: True
tag optional	String	Filter segments by tag(s)
operation_match optional	String	Operation to match. If None only editdistance is calculated Default value: regex,tags,posTag,split Allowed values: 'regex,tags,posTag,split', 'None'

Success 200

Field	Type	Description
-------	------	-------------

Translation	String/ Json	units matching the query
-------------	-----------------	--------------------------

Error 4xx

Name	Type	Description
401	String	Unauthorized
403	String	Insufficient permissions

2.2.1.2 Batch search translation memory segments

GET -> https://localhost/api/v1/tm/query_batch

Permission: user

Field	Type	Description
token	String	Token returned by auth endpoint.

Parameter

Field	Type	Description
q	String	String to query (multiple values allowed)
slang	String	Source language.

tlang	String	Target language.
smeta optional	String	Source metadata (stringified valid JSON). For example: <i>JSON.stringify({"x-context-post": "pretag", "x-context-pre": "posttag"})</i>
tmeta optional	String	Target metadata (stringified valid JSON)
limit optional	Number	Limit output to this number of segments Default value: 10
min_match optional	Number	Return only match above or equal to given threshold (0-100) Default value: 75
strip_tags optional	Boolean	Strip all XML tags from the query Default value: false
concordance optional	Boolean	Concordance search mode Default value: false

aut_trans optional	Boolean	Apply machine translation if match score is less than a threshold Default value: True
tag optional	String	Filter segments by tag(s)
operation_match optional	String	Operation to match. If None only editdistance is calculated Default value: regex, tags, posTag, split Allowed values: 'regex, tags, posTag, split', 'None'
file_name optional	String	Filter segments by filename(s).
organization optional	String	Filter segments by organization(s).
industry optional	String	Filter segments by industry(s).
language optional	String	Filter segments by language(s).
type optional	String	Filter segments by type(s).
username optional	String	Filter segments by creator's username.
tm_change_date optional	Date	Filter segments by TM change date (.from and .to limits).

tm_creation_date optional	Date	Filter segments by TM creation date (.from and .to limits).
insert_date optional	Date	Filter segments by date of insertion into DB (.from and .to limits).
update_date optional	Date	Filter segments by date of update in DB (.from and .to limits).
check_date optional	Date	Filter segments by script check (.from and .to limits).
dirty_score optional	Integer	Filter segments by clean score (.from and .to limits).

Filter Example

```
{
  "tm_change_date.from": "20160414",
  "tm_change_date.to": "20160822",
  "check_date.to": "20160530"
}
```

Success 200

Field	Type	Description
Translation	String/Json	units matching the query

Error 4xx

Name	Type	Description
401	String	Unauthorized
403	String	Insufficient permissions

2.2.2.6 Authorization

POST -> <https://localhost/api/v1/auth>

Parameter

Field	Type	Description
username	String	Name of User
password	String	Password for account

Error 4xx

Name	Type	Description
401	String	Invalid credentials

Success 200

Field	Type	Description
-------	------	-------------

access_token	String	Authorization token for use it other endpoints
--------------	--------	--

2.2.3.3 Get tag details

GET -> <https://localhost/api/v1/tags/<id>>

Permission: user

Field	Type	Description
token	String	Token returned by auth endpoint.

Parameter

Field	Type	Description
tag optional	String	Name of Tag, if not Tag specified, all tag' details are returned. Users will only see allowed tags in their scopes.

Error 4xx

Name	Type	Description
404	String	Tag doesn't exist
403	String	Insufficient permissions

401	String	Unauthorized
-----	--------	--------------

2.2.4 Import Services

Services related to the import of content in NECTM. The data to be imported can be either:

- TMX files or
- Single translation memory unit

It's possible to specify tags for categorization of the content (refer to section [2.1.3](#)).

2.2.4.1 Import translation memory segments from TMX file

PUT -> <https://localhost/api/v1/tm/import>

Permission: admin, users authorized

Field	Type	Description
token	String	Token returned by auth endpoint.

Parameter

Field	Type	Description
file	File	Zipped TMX file to import.
tag	String	Tag name of the imported file.
lang_pair optional	String	Language pair to import (for multilingual TMX files). 2-letter language codes join by underscore. By default, import first pair in each segment

Error 4xx

Name	Type	Description
401	String	Unauthorized
403	String	Insufficient permissions

2.2.4.2 Add new translation memory unit

POST -> <https://localhost/api/v1/tm>

Permission: users authorized

Field	Type	Description
token	String	Token returned by auth endpoint.

Parameter

Field	Type	Description
s-text	String	Source segment text.
t-text	String	Target segment text..
s-lang	String	Source language.
t-lang	String	Target language.

tag	String	Translation unit tag.
file_name optional	String	File name (or source name)

Error 4xx

Name	Type	Description
401	String	Unauthorized
403	String	Insufficient permissions

2.2.5 Export Services

2.2.5.1 Export translation memory segments to zipped TMX file(s)

<https://localhost/api/v1/tm/export>

Permission: admin, users authorized

Field	Type	Description
token	String	Token returned by auth endpoint.

Parameter

Field	Type	Description
slang	String	Source language.

tlang	String	Target language.
squery optional	String	Filter source segments by this query. Supports regular expressions.
tquery optional	String	Filter target segments by this query. Supports regular expressions.
limit optional	Number	Limit number of exported segments.
duplicates_only optional	Boolean	Return duplicate segments only (duplicates = having the same source text)
file_name optional	String	Filter segments by filename(s).
organization optional	String	Filter segments by organization(s).
tag optional	String	Filter segments by tag(s).
industry optional	String	Filter segments by industry(s).
language optional	String	Filter segments by language(s).
type optional	String	Filter segments by type(s).
username optional	String	Filter segments by creator's username.

tm_change_date optional	Date	Filter segments by TM change date (.from and .to limits).
tm_creation_date optional	Date	Filter segments by TM creation date (.from and .to limits).
insert_date optional	Date	Filter segments by date of insertion into DB (.from and .to limits).
update_date optional	Date	Filter segments by date of update in DB (.from and .to limits).
check_date optional	Date	Filter segments by script check (.from and .to limits).
dirty_score optional	Integer	Filter segments by clean score (.from and .to limits).

Success 200

Field	Type	Description
task_id	String	ID of export task invoked in the background

Error 4xx

Name	Type	Description
------	------	-------------

401	String	Unauthorized
403	String	Insufficient permissions

2.2.5.2 Download exported file or list all available downloads

https://localhost/api/v1/tm/export/file/<export_id>

Permission: admin, users authorized

Field	Type	Description
token	String	Token returned by auth endpoint.

Success 200

Field	Type	Description
binary	File	Content of zipped TMX file(s) (if export_id is supplied)
files	Json	List of all available exports (if export_id is not supplied)

Error 4xx

Name	Type	Description
------	------	-------------

401	String	Unauthorized
403	String	Insufficient permissions