Traducción automática estadística para traductores

Pilar Sánchez Gijón (UAB)

Antoni Oliver (UOC)

Sede de la Comisión Europea

Madrid, 25 de junio de 2018

Objetivo

Entender y gestionar un sistema de traducción automática estadística basado en Moses.

1. ¿En qué consiste un sistema de TAE?
2. ¿Qué opciones tengo?
3. Si decido tener mi propio sistema, ¿cómo puedo hacerlo?

1.¿En qué consiste un sistema de TAE?

A partir de una gran cantidad de texto en la lengua de partida y en la lengua de llegada, alineado y no alineado, el sistema busca la mejor traducción posible de cada uno de los elementos de una oración, partiendo del segmento más grande posible (frase completa) hasta el más pequeño (palabra).

1.¿En qué consiste un sistema de TAE?

La calidad de la traducción depende de:

1. El corpus de trabajo:
   - Córpora monolingües y corpus bilingüe

2. El tiempo de respuesta:
   - Cantidad de opciones valoradas
   - Cantidad de texto procesado

2. ¿Qué opciones tengo?

- Sistemas de TA (TAE u otros sistemas) ya existentes
- Crear un motor en un sistema público
- Construirse un sistema propio

2. ¿Qué opciones tengo?

Sistemas de TA (TAE u otros sistemas) ya existentes:

VENTAJAS:

- Solución tecnológica integral
- No hay necesidad de buscar/gestionar corpus

2. ¿Qué opciones tengo?

Sistemas de TA (TAE u otros sistemas) ya existentes:

INCONVENIENTES:

- Seguridad
  ¿Dónde quedan los archivos que traduzco?
- Coste
  ¿Es un SAAS? ¿Coste por volumen de palabras?
- Interacción con herramientas TAO
  ¿Tiene API o se integra de cualquier otro modo?
- Adaptabilidad:
  ¿El corpus prevé los tipos de texto que me interesa traducir?

2. ¿Qué opciones tengo?

Crear un motor en un sistema público:

VENTAJAS:

- Solución tecnológica integral
- Adaptabilidad del corpus
- Seguridad: acceso restringido a archivos y córpora

2. ¿Qué opciones tengo?

Sistemas de TA (TAE u otros sistemas) ya existentes:

INCONVENIENTES:

- Seguridad
  ¿Dónde quedan los archivos que traduzco?
- Coste
  ¿Es un SAAS? ¿Coste por volumen de palabras?
- Interacción con herramientas TAO
  ¿Tiene API o se integra de cualquier otro modo?
- Gestión de archivos:
  ¿Con qué archivos trabajo? ¿Dónde los obtengo? ¿Cómo los gestiono?

2. ¿Qué opciones tengo?

Construirse un sistema propio

VENTAJAS:

- Adaptabilidad del corpus
- Seguridad: acceso restringido a archivos y córpora

2. ¿Qué opciones tengo?

Construirse un sistema propio:

INCONVENIENTES:

- Tecnología
  ¿Programación? ¿Linux? ¿Gestión de un servidor?
- Coste
  ¿Servidor? ¿Aspectos de seguridad?
- Interacción con herramientas TAO
  ¿Con qué TAO? ¿Cómo integrarlo
- Gestión de archivos:
  ¿Con qué archivos trabajo? ¿Dónde los obtengo? ¿Cómo los gestiono?

3. Si decido tener mi propio sistema, ¿cómo puedo hacerlo?

MTradumàtica

Approprite to start managing and testing MT technologies.

Intuitive creation and management of SMT engines in five steps:

5 Steps

(1) Upload files
(2) Create and manage monotexts
(3) Build language models (LMs)
(4) Create and manage bitexts
(5) Train SMT translation models.

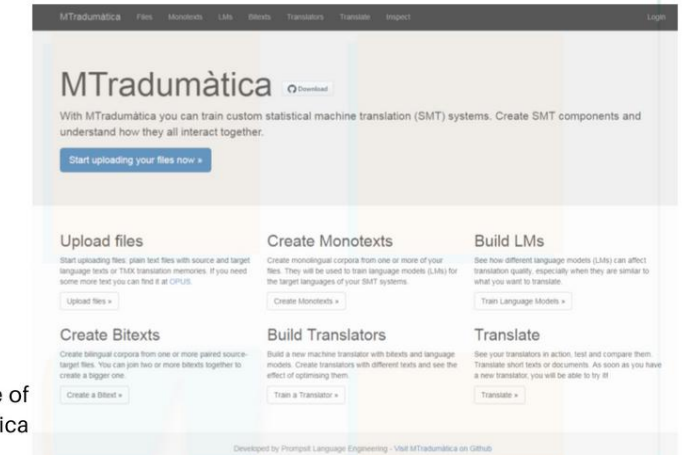
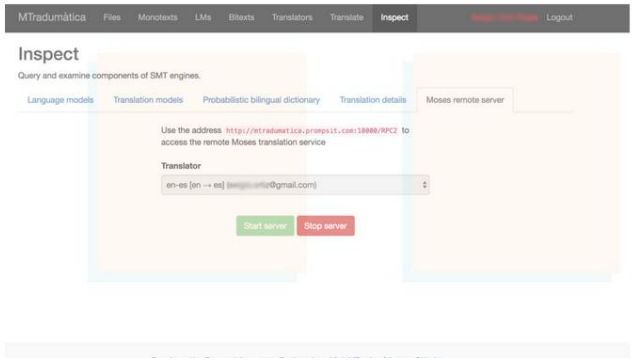


Figure 1. The interface of MTradumàtica

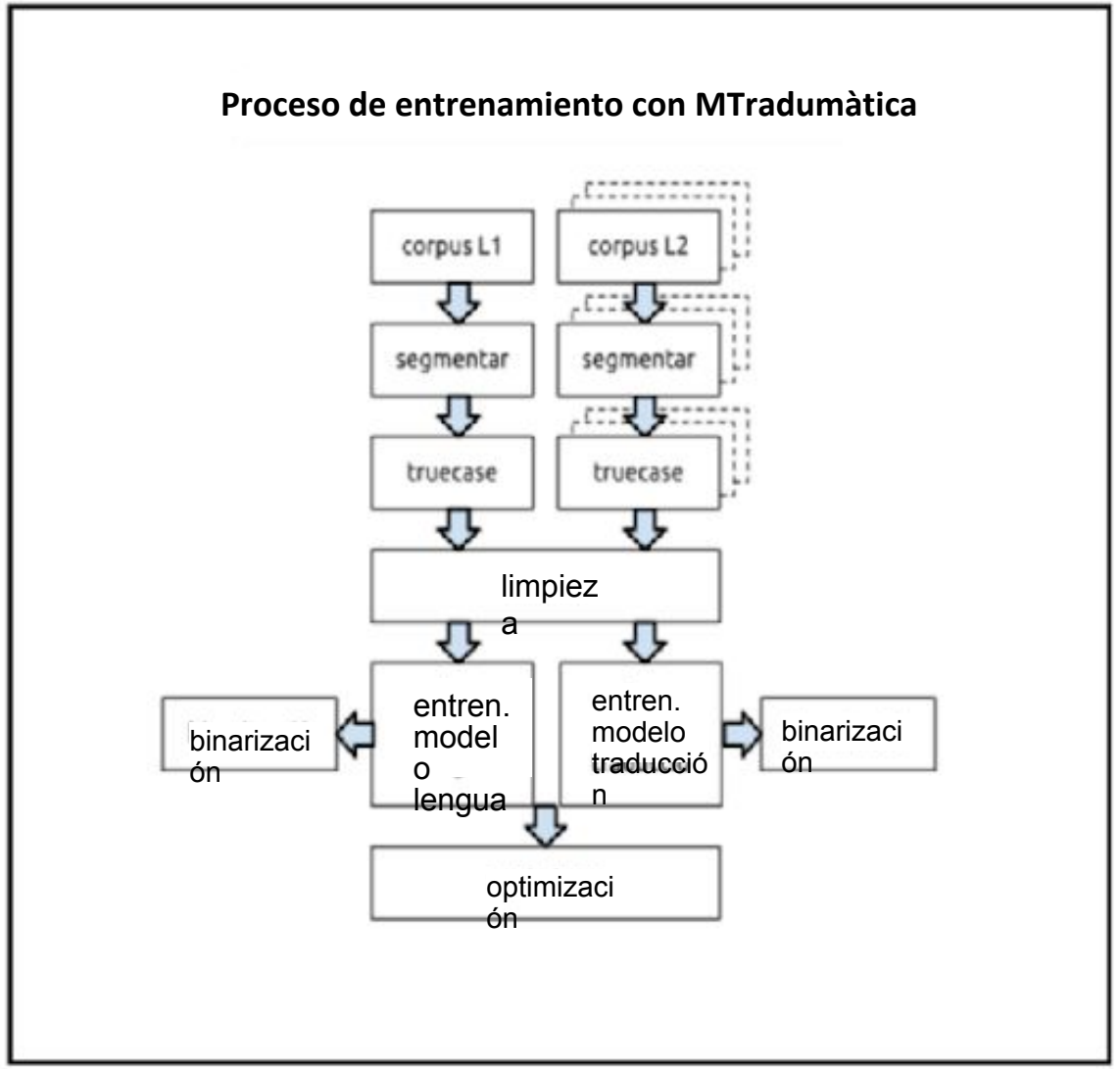
- Works with parallel, sentence-by-line text files such as Moses.
- Has a user-friendly web interface.
- Can be installed in private servers (increased confidentiality).

- Is able to import TMX files.
- Includes user management features (increased confidentiality).
- Integration with TM systems, like OmegaT, through an URL.




OmegaT 4.1.3_1 :: prova
Projecte Edita Vés a Visualització Eines Opcions Ajuda
"Inglés" ;
"Switch to English";
<segment 2384>
"Contacts";
"Calls";
1. "English";
"Inglés" ;
<50/33/50% >
" Cambiar a inglés " ;
<Moses>
Diccionari
Diverses traduccions
Notes
Comentaris
Glossari
Segment Properties

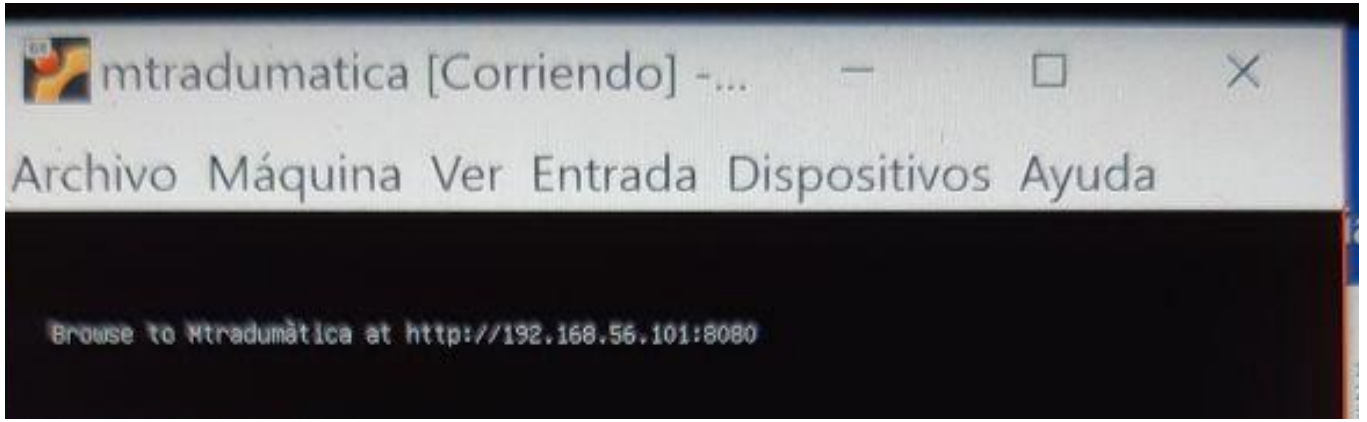
3.1 Proceso de entrenamiento de un motor TAE



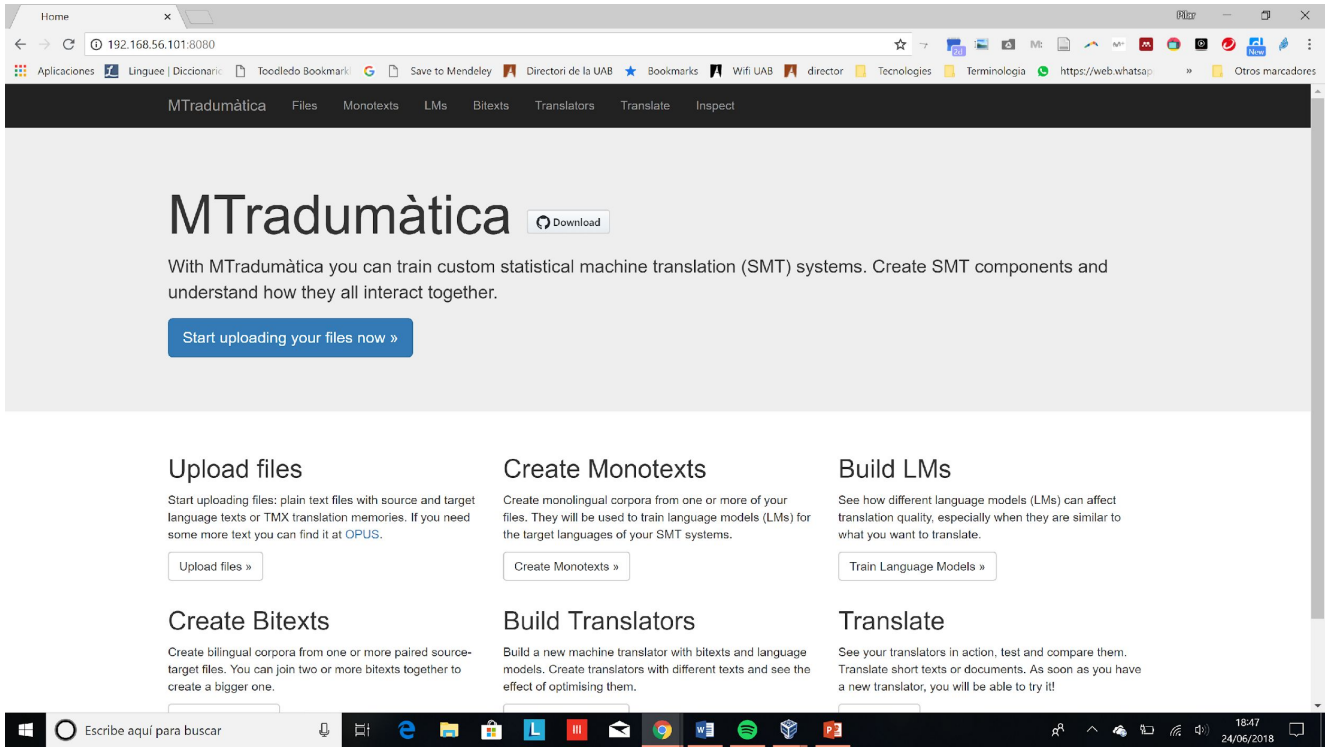
3.2. Proceso de instalación

1. Instalar Virtualbox
2. Importar Mtradumatica.ova
3. Cambiar la configuración del motor
4. Iniciar el motor
5. Generar la URL para trabajar a través del navegador

3.2 URL de acceso al servidor en local

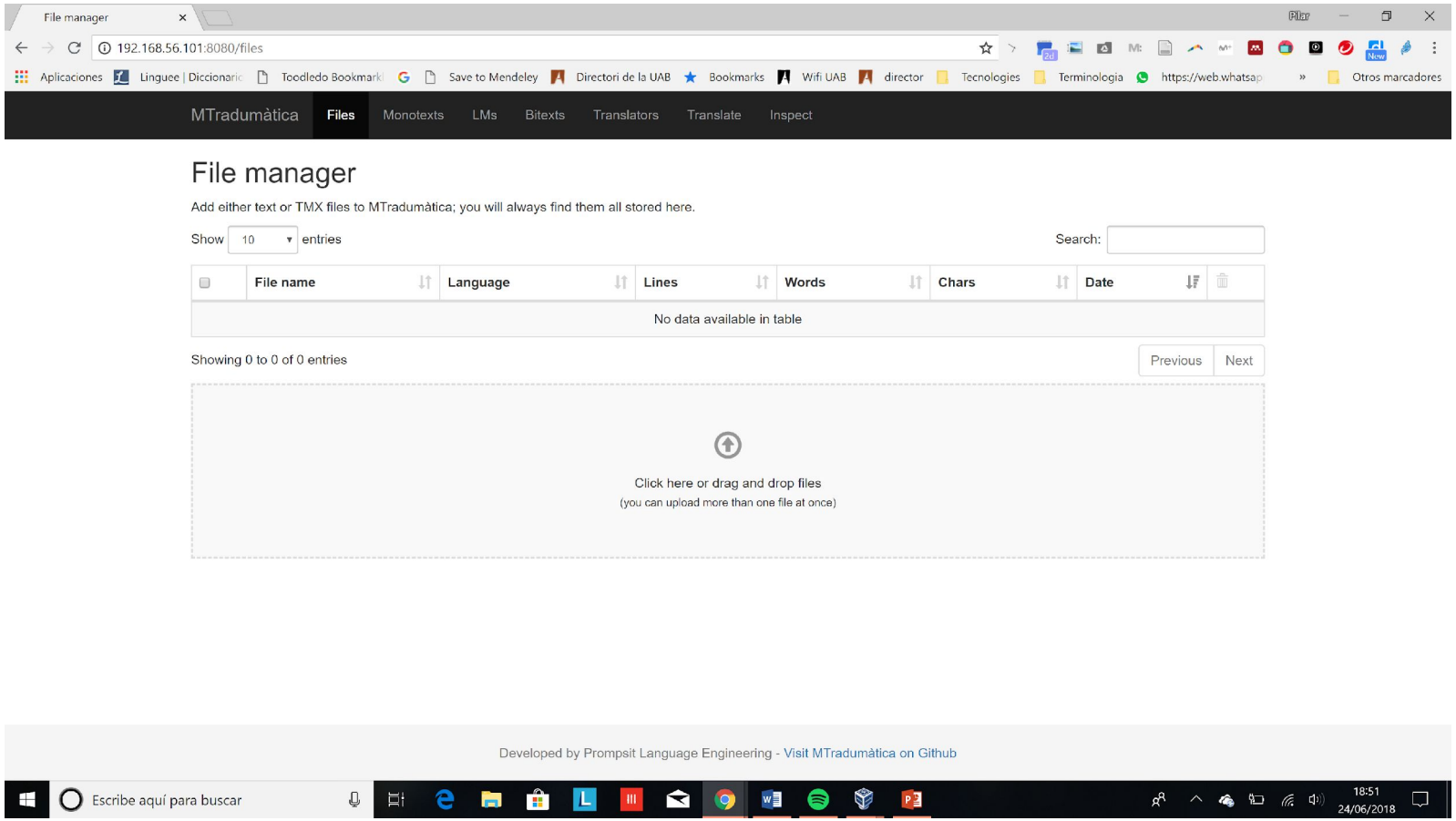


Proceso



1. Subir archivos monolingües o bilingües
2. Crear monotextos
3. Construir modelos de lengua (LM)
4. Crear bitextos
5. Construir traductores (motores)
6. Traducir

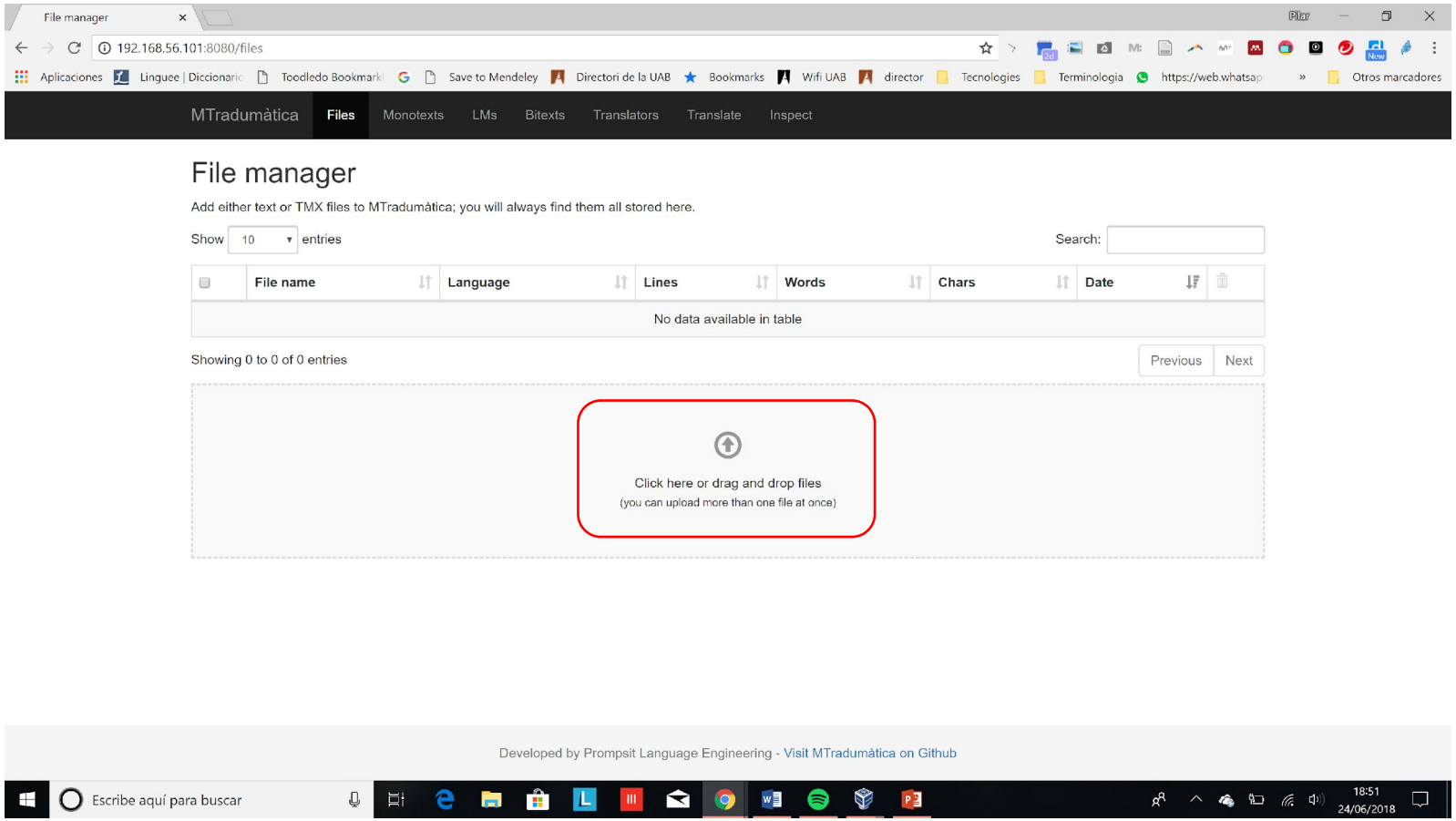
Upload files



Formatos de archivo:

- Txt (monolingüe)
- Txt (alineado para Moses)
- TMX

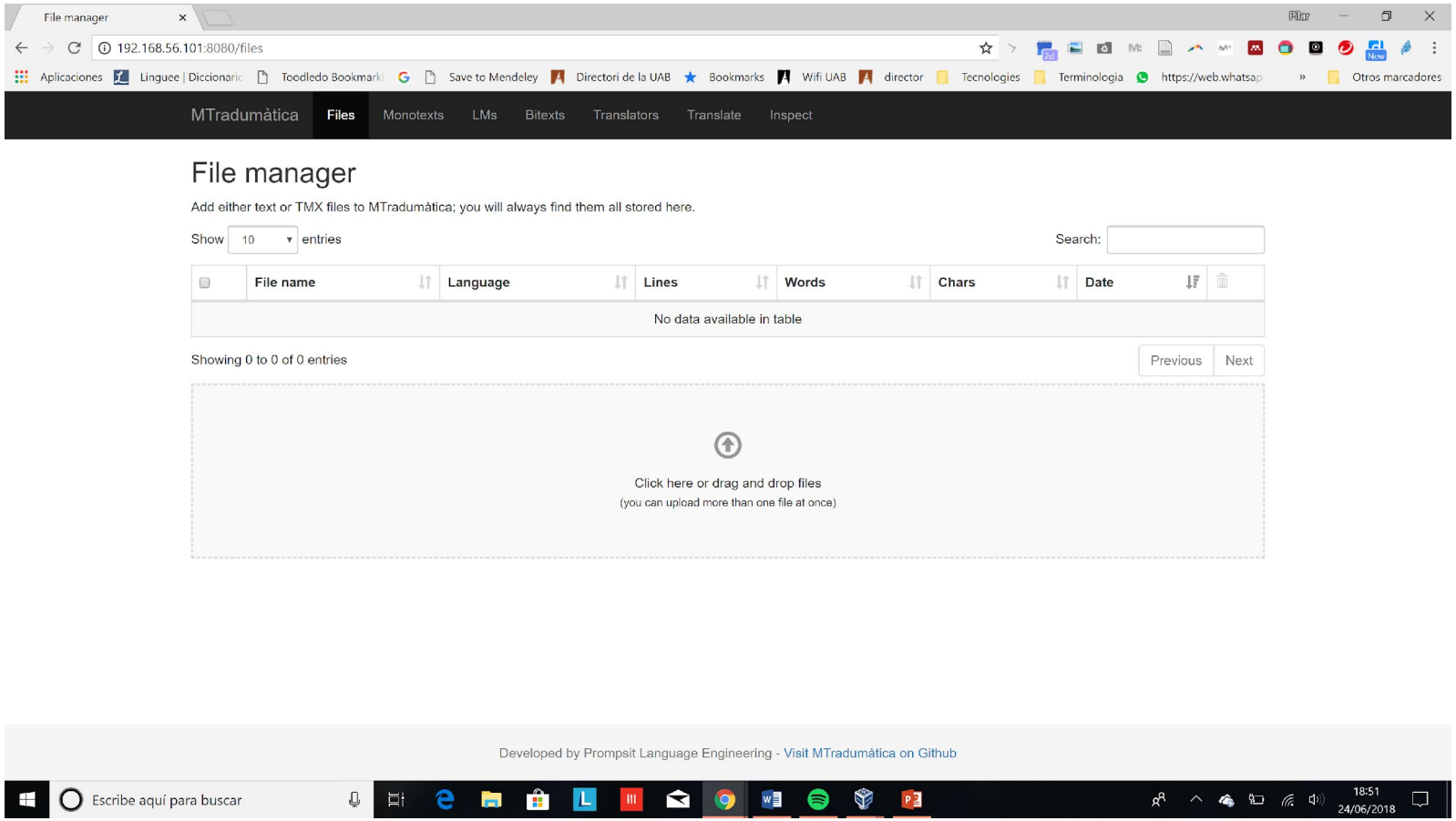
Upload files



Formatos de archivo:

- Txt (monolingüe)
- Txt (alineado para Moses)
- TMX

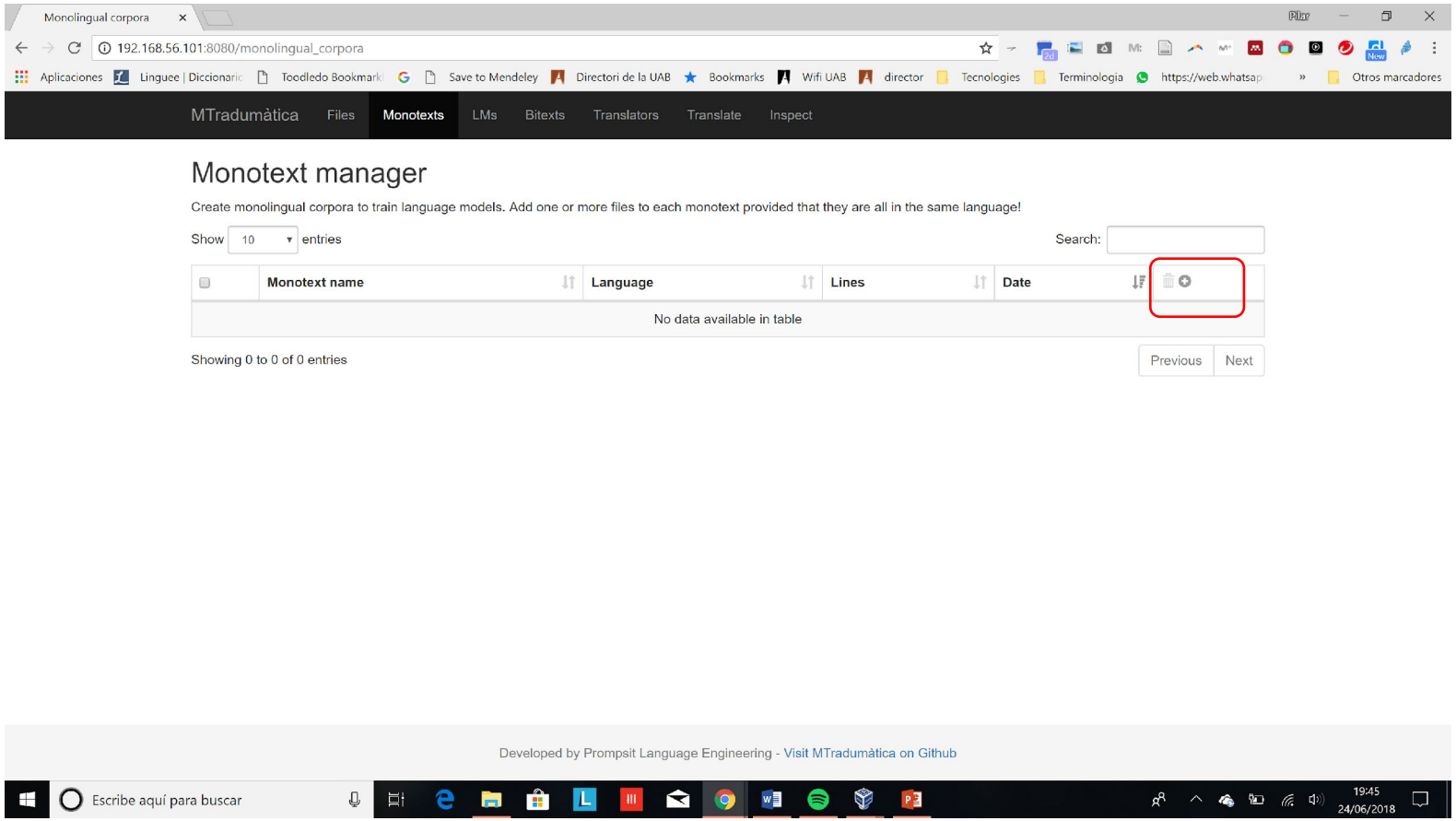
Upload files



Formatos de archivo:

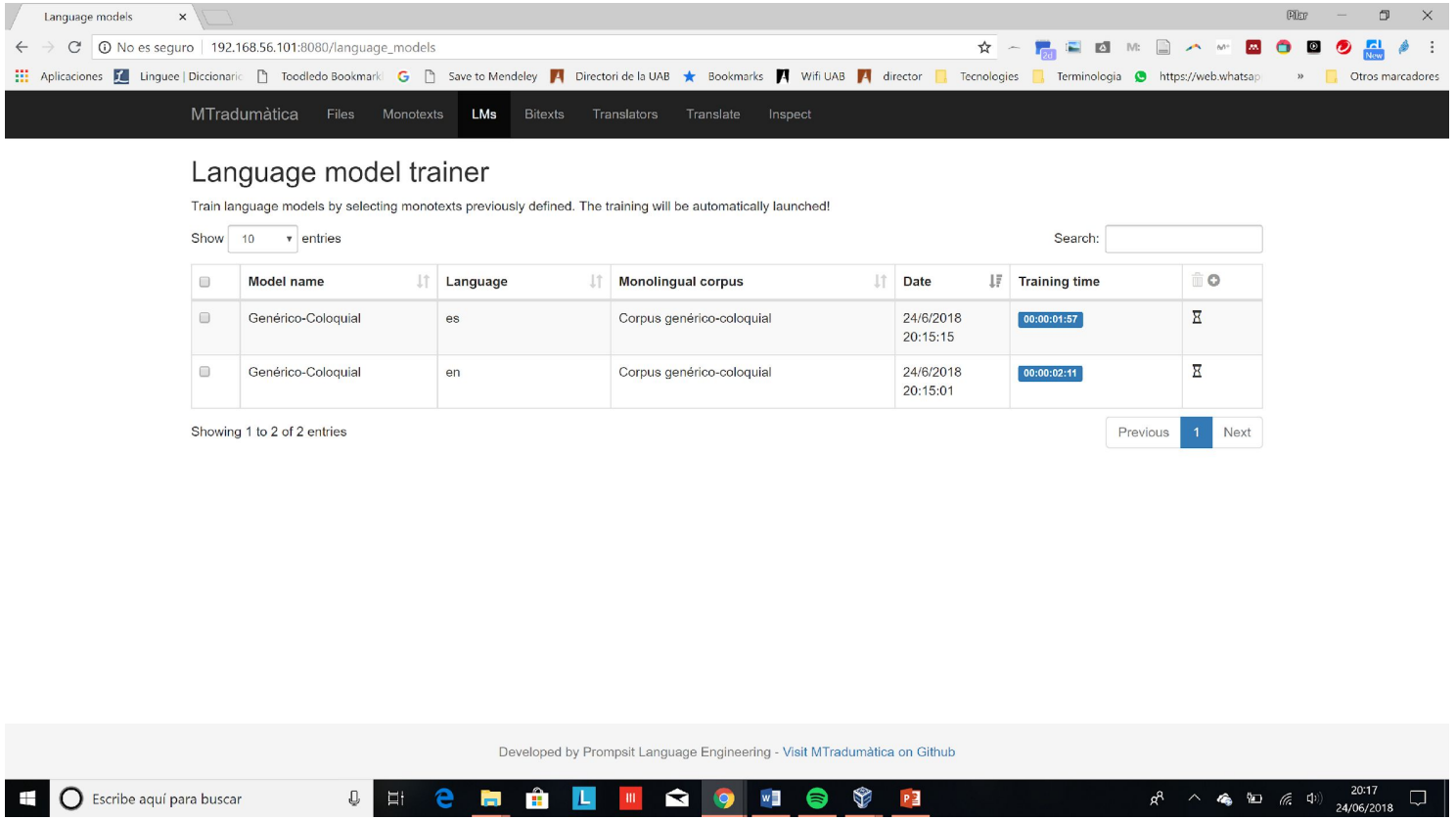
- Txt (monolingüe)
- Txt (alineado para Moses)
- TMX

Monotext Manager



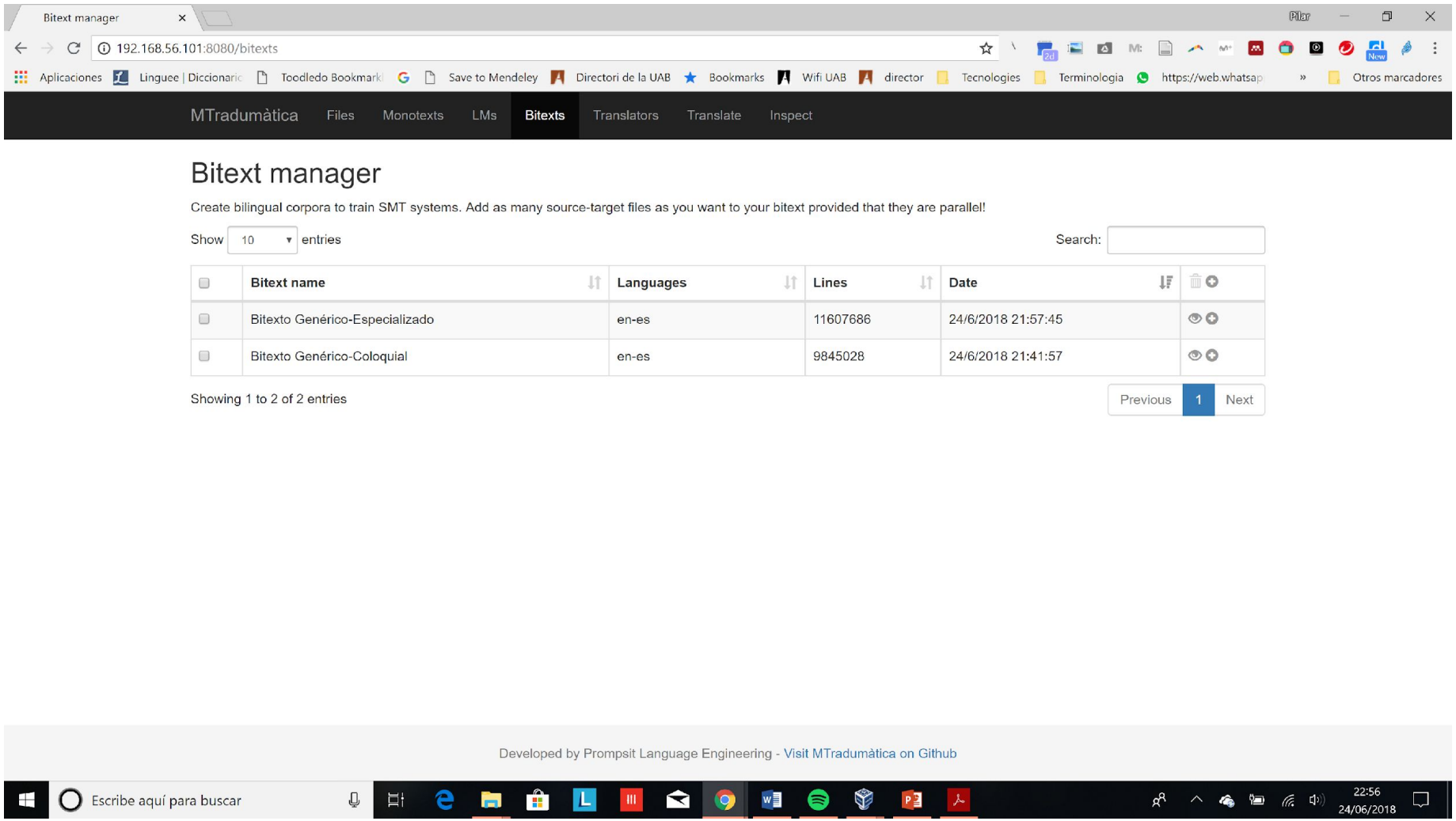
Crear corpus monolingües con texto para cada combinación lingüística. Puede ser corpus que después también se alinea.

Language Model trainer

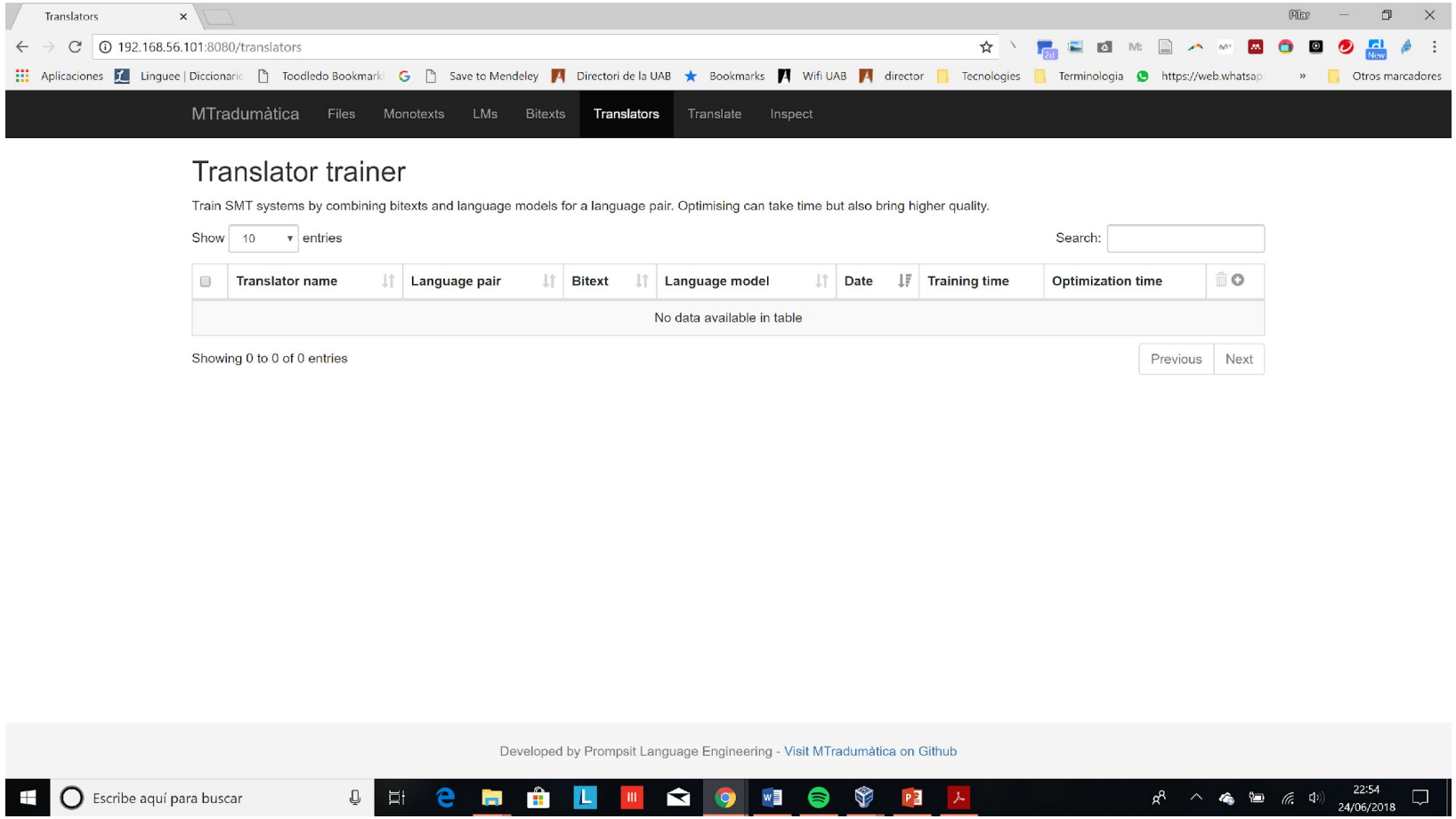


- Entrenamiento de cada modelo de lengua a partir de un monotexto. Consiste en binarizar los textos.
- Se puede entrenar un modelo de lengua por monotexto.
- El entrenamiento lleva su tiempo, dependiendo del tamaño del archivo.

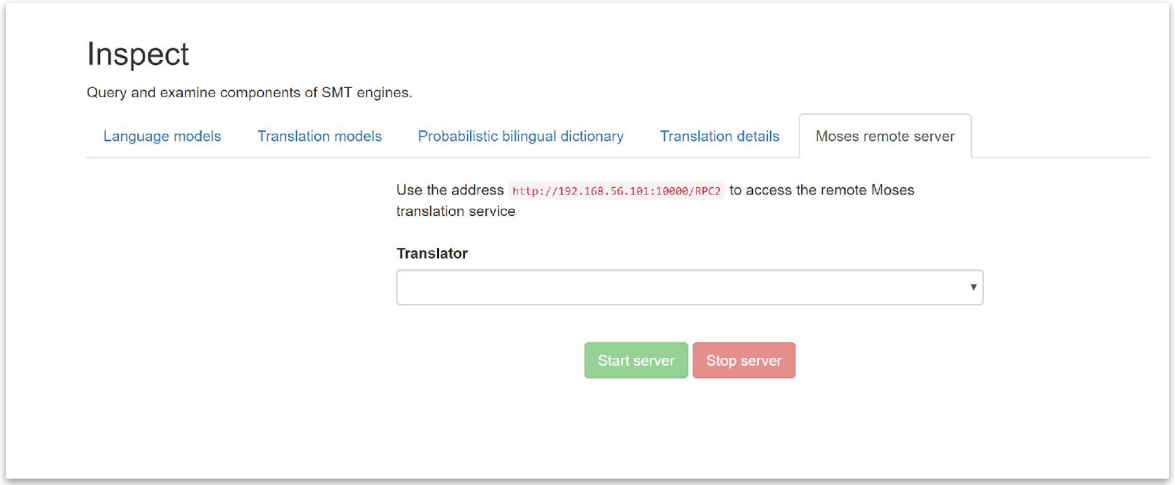
Bitext Builder



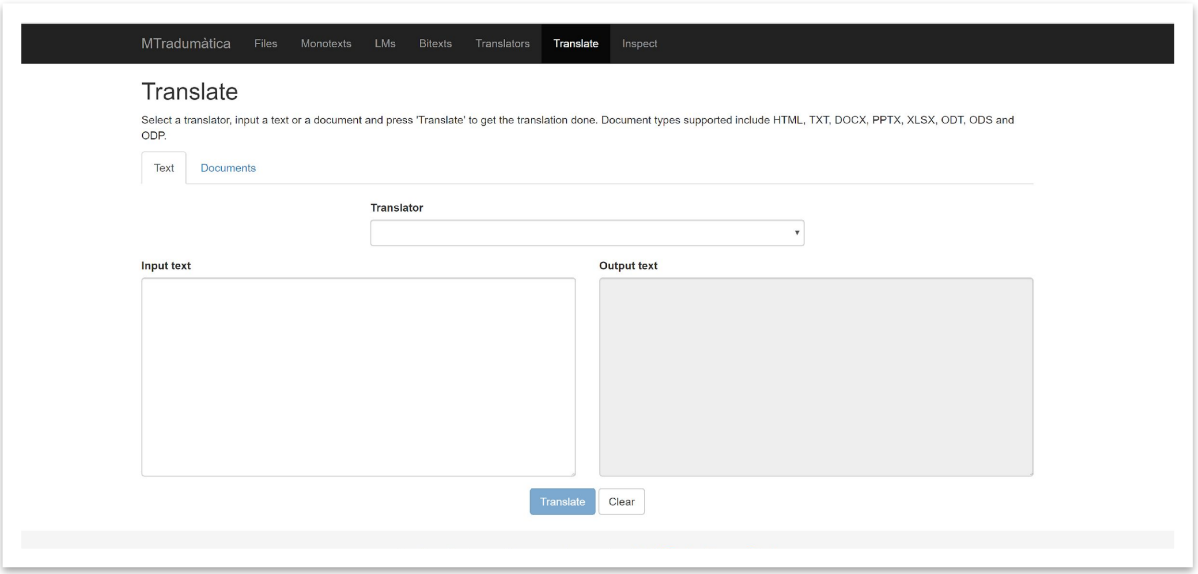
Translator trainer



.....y traduce!



Vincula el motor con OmegaT mediante esta URL.



Traducir textos o frases.

¿Lo intentamos?

- En tu ordenador:
  - Instala Virtualbox
  - Importa mtradumatica.ova
  - Ajusta la configuración (RAM y núcleos)
  - Inicia la máquina
  - ….. y comienza a gestionar tu sistema de motores!!!

Referencias bibliográficas

MARTÍN-MOR, A.; PIQUÉ, R. (2017). «MTradumàtica i la formació de traductors en Traducció Automàtica Estadística». Revista Tradumàtica. Tecnologies de la Traducció, 15, 97-115. https://doi.org/10.5565/rev/tradumatica.199