

Desarrollo de Software Libre: Caso Esquite

Diego A. Barriga (@umoqnier)

21 de Octubre 2020

Elotl / LIDSOL

\$ whoami



- LIDSOL
 - ¿Software Libre?
 - Libertades
- Comunidad Elotl
 - Diversidad lingüística
- Esquite
 - ¿Qué es?
 - ¿Qué hace?
 - ¿Cómo contribuir?
- Otros proyectos
 - Glosador
 - Axolotl
 - `import elotl`



Laboratorio de Investigación y Desarrollo
de Software Libre

Figure 2: LIDSOL



Figure 3: Equipo de LIDSOL

¿Con qué se come el software libre?

- El software libre es un movimiento social

¿Con qué se come el software libre?

- El software libre es un movimiento social
- Miles de voluntarios creando software

¿Con qué se come el software libre?

- El software libre es un movimiento social
- Miles de voluntarios creando software
- Modelo de producción de conocimiento

¿Con qué se come el software libre?

- El software libre es un movimiento social
- Miles de voluntarios creando software
- Modelo de producción de conocimiento
- Esquema justo de licenciamiento



Figure 4: Mascota del proyecto GNU

¿De a cómo la libertad?

Para que un bicho sea considerado software libre debe cumplir 4 libertades que se muestran a continuación:

- **Libertad de uso**
 - Podemos usarlo para lo que queramos
- **Libertad de aprendizaje**
 - Podemos aprender cómo es que está hecho y además tenemos acceso al código fuente
- **Libertad de modificación**
 - Podemos adecuar el programa a nuestras necesidades, corregir defectos, incorporar pedazos del programa a nuestros proyectos
- **Libertad de distribución**
 - Podemos compartir el programa con otras personas, tal como lo recibimos o con nuestras modificaciones

¿Pa' qué?

- El software es *conocimiento*
- El *conocimiento* es información (y poder)
- La información es **ciencia**

El software es ciencia



Figure 5: Conocimiento es poder

Desarrollos en México



Figure 6: mapita de México

Desarrollos en México



Figure 6: mapita de México

- ¿Cuántas lenguas se hablan en México?

Desarrollos en México



Figure 6: mapita de México

- ¿Cuántas lenguas se hablan en México?
- ¿Hay un traductor para estas lenguas o una aplicación?

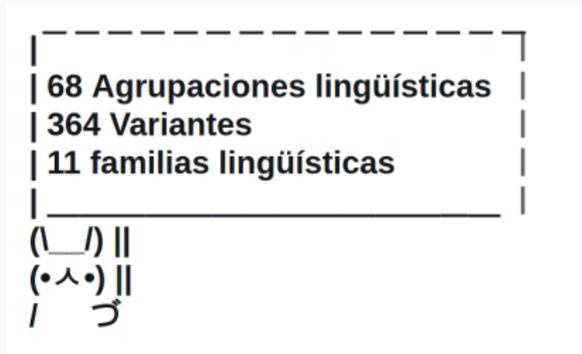


Figure 7: Diversidad lingüística

- La investigación y desarrollo para lenguas mexicanas no solo tiene un **impacto social** positivo
- También supone un **reto tecnológico y 100tífico** por las características de estas lenguas



Figure 8: Un gato

- Casi no hay desarrollos tecnológicos para este tipo de lenguas
 - Son lenguas de escasos recursos digitales
 - Falta de interés o desconocimiento
 - Gran variedad dialectal y falta de normalización de las lenguas

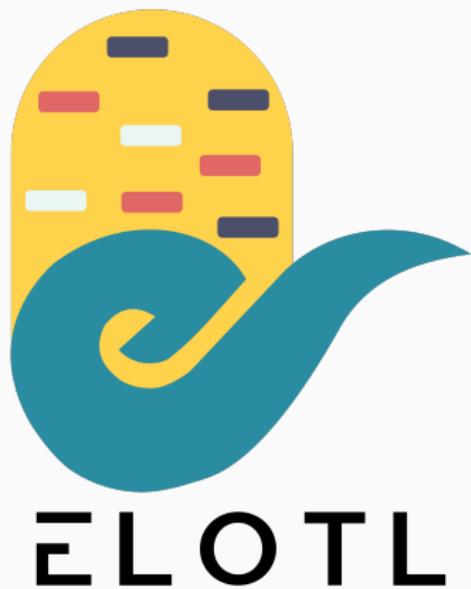


Figure 9: Un elotito :3

¿Qué es Elotl?

- Grupo de voluntarias, becarias, académicas e interesadas en el desarrollo de **tecnologías del lenguaje** para lenguas habladas en México

¿Qué es Elotl?

- Grupo de voluntarias, becarias, académicas e interesadas en el desarrollo de **tecnologías del lenguaje** para lenguas habladas en México
- Herramientas **libres** y **gratuitas** <3

¿Qué es Elotl?

- Grupo de voluntarias, becarias, académicas e interesadas en el desarrollo de **tecnologías del lenguaje** para lenguas habladas en México
- Herramientas **libres** y **gratuitas** <3
- **Difusión** de cuestiones lingüísticas, sociales y tecnológicas de lenguas mexicanas

¿Qué es Elotl?

- Grupo de voluntarias, becarias, académicas e interesadas en el desarrollo de **tecnologías del lenguaje** para lenguas habladas en México
- Herramientas **libres** y **gratuitas** <3
- **Difusión** de cuestiones lingüísticas, sociales y tecnológicas de lenguas mexicanas
- Comunidad completamente virtual (desde antes de la plaga)

Esquite



Ejemplo de implementación: TSUNKUA

CORPUS PARALELO
OTOMÍ-ESPAÑOL

TSUNKUA



2019 | INTERNATIONAL YEAR OF
Indigenous Languages

TSUNKUA

Ayuda  Ligas de interés  Acerca de 

Consulta del Corpus

Idioma

Español

Búsqueda

Palabra o frase a buscar



Resultados por
variante
(opcional)

Selecciona una o más variantes

¿Qué es Esquite?

- *Framework* para administrar corpus paralelos (textos bilingües)

¿Qué es Esquite?

- *Framework* para administrar corpus paralelos (textos bilingües)
- Sistema web para hacer búsquedas avanzadas por palabras o frases

¿Qué es Esquite?

- *Framework* para administrar corpus paralelos (textos bilingües)
- Sistema web para hacer búsquedas avanzadas por palabras o frases
- Administrador de documentos paralelos

¿Qué es Esquite?

- *Framework* para administrar corpus paralelos (textos bilingües)
- Sistema web para hacer búsquedas avanzadas por palabras o frases
- Administrador de documentos paralelos
- API para hacer búsquedas de forma programática



Figure 10: python / django



Figure 11: elasticsearch

Administrador de corpus

Administrador del Corpus

6 documentos  | 5515 líneas totales  | 3 Variantes  | [Exportar](#) 

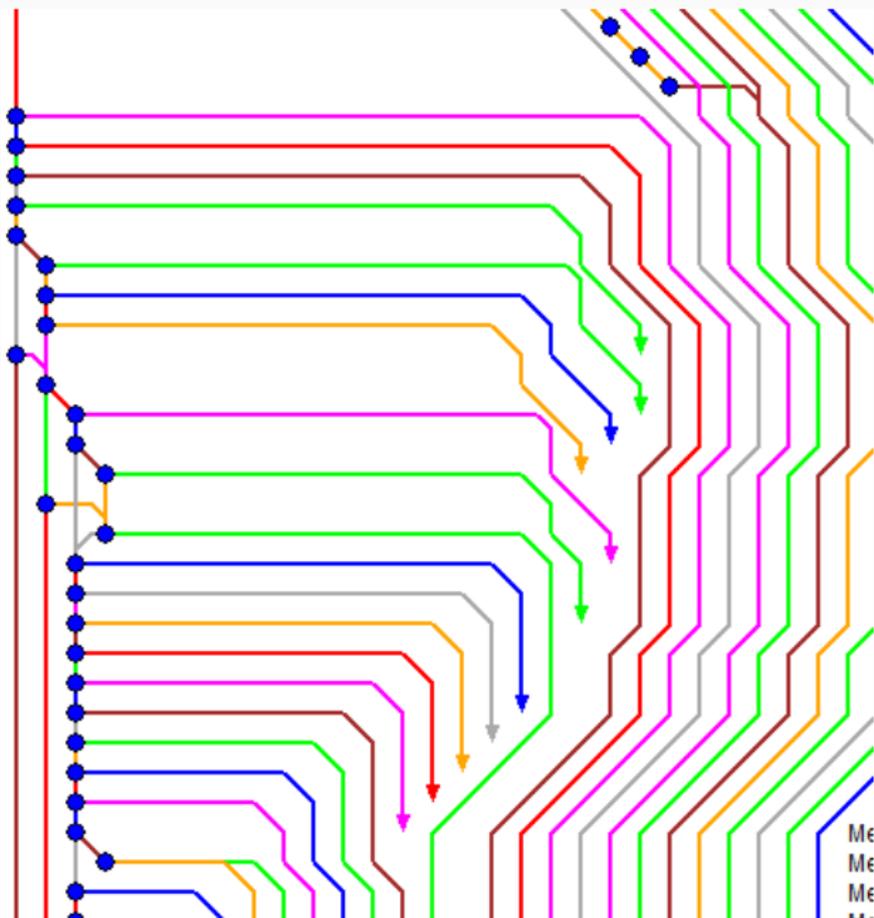
Nombre	Archivo	Párrafos	Acciones
Visión de los vencidos (hñahñu)	visiondelosvencidoshñahnu.pdf	3225	Ver  Editar  Añadir  Eliminar 
El otomí de toluca, Yolanda Lastra	elotomidetoluca.pdf	1638	Ver  Editar  Añadir  Eliminar 
NORMA DE ESCRITURA DE LA LENGUA HÑĀHÑU (OTOMÍ)	Norma_Otomí.pdf	404	Ver  Editar  Añadir  Eliminar 
Notas gramaticales, Doris Bartholomew	ote_gramsel.pdf	221	Ver  Editar  Añadir  Eliminar 
PRONTUARIO EN OTOMÍ / ESPAÑOL HÑĀHÑU	p_otomí_f.pdf	15	Ver  Editar  Añadir  Eliminar 
Himno Nacional Mexicano en Lengua Otomí (INPI)	Documento_digital_no_disponible.pdf	12	Ver  Editar  Añadir  Eliminar 

[Nuevo Documento](#) 

ESQUITE DEV 2020 |  por [Esquite](#)  | Hecho con  por [Comunidad Eloti](#)

Figure 12: Administrador de corpus

¿Cómo puedo contribuir?



¿Te gusta programar?

- Con código, *features* y mejoras

¿Te gusta programar?

- Con código, *features* y mejoras
- Visita nuestro repositorio
 - CONTRIBUTING
 - Docs

¿Te gusta programar?

- Con código, *features* y mejoras
- Visita nuestro repositorio
 - CONTRIBUTING
 - Docs
- ¿Aun tienes dudas de como mandar tus mejoras?
 - Guías para contribuir
 - Contactanos :)
 - contacto at elotl.mx
 - diegobarriga at protonmail.com
 - @elotlmx

No sabo programar ¿Cómo está el bisnes?

- Usa el *framework*

No sabo programar ¿Cómo está el bisnes?

- Usa el *framework*
- Sugiere mejoras, características nuevas o cambios en la documentación

No sabo programar ¿Cómo está el bisnes?

- Usa el *framework*
- Sugiere mejoras, características nuevas o cambios en la documentación
- Recopilación de textos bilingües
 - identificación
 - digitalización
 - procesamiento

No sabo programar ¿Cómo está el bisnes?

- Usa el *framework*
- Sugiere mejoras, características nuevas o cambios en la documentación
- Recopilación de textos bilingües
 - identificación
 - digitalización
 - procesamiento
- Difusión
 - Recomendanos con quien mas confianza le tengas

No sabo programar ¿Cómo está el bisnes?

- Usa el *framework*
- Sugiere mejoras, características nuevas o cambios en la documentación
- Recopilación de textos bilingües
 - identificación
 - digitalización
 - procesamiento
- Difusión
 - Recomendanos con quien mas confianza le tengas
- Investigación y desarrollo (NLP + *low resources*)
 - Integrantes de Elotl pueden dirigir tesis
 - En LIDSOL podrías liberar tu servicio social
 - Si eres profesxr podríamos ponerte en contacto con entusiastas del área

No sabo programar ¿Cómo está el bisnes?

- Usa el *framework*
- Sugiere mejoras, características nuevas o cambios en la documentación
- Recopilación de textos bilingües
 - identificación
 - digitalización
 - procesamiento
- Difusión
 - Recomendanos con quien mas confianza le tengas
- Investigación y desarrollo (NLP + *low resources*)
 - Integrantes de Elotl pueden dirigir tesis
 - En LIDSOL podrías liberar tu servicio social
 - Si eres profesxr podríamos ponerte en contacto con entusiastas del área

```
import elotl
print("Name\t\tDescription")
list_of_corpus = elotl.corpus.list_of_corpus()
for row in list_of_corpus:
    print(row)
```

```
Name          Description
['axolotl', 'Is a Spanish-Nahuatl parallel corpus']
['tsunkua', 'Is a Spanish-otomí parallel corpus']
```

Seleccionar idioma ▾

Con la tecnología de  Traductor de Google



Inicio Buscar ▾ Ayuda Ligas de interés

Búsquedas AXOLOTL

Idioma

Búsqueda

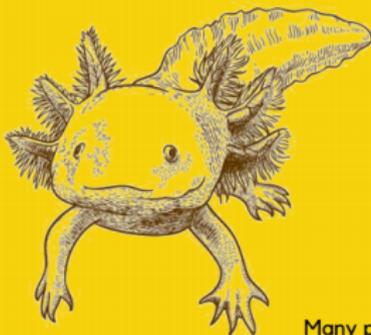
[¿Cómo hacer búsquedas avanzadas?](#)

[Términos de uso](#) | [¿Cómo citar AXOLOTL?](#)
2020 México, D.F.

Corpus Paralelo Axolotl

Ammo ticmatih intla axolotl
noyuh salamandra
nahuatlahtolcopa itoca aitzcuin

El ajolote no saben
muchos que es
salamandra en la lengua
mexicana le llaman
"perro del agua"



Many people do not know
the ajolote is a salamander in the
ancient Mexican tongue they
called it "dog of the waters".

Contacto

- *diegobarriga at protonmail*
- *contacto at elotl.mx*
- @elotl_ - Twitter
- elotl.mx
- `github.com/ElotlMX`

Slides

- `github.com/ElotlMX/ccross-2020-esquite`

- Elotl
 - Repo Esquite
 - Repo modulo elotl
 - Tsunkua
 - Axolotl
 - Materiales didácticos
- LIDSOL
- Guías de contribución