Clinical language model on Mexican medical records
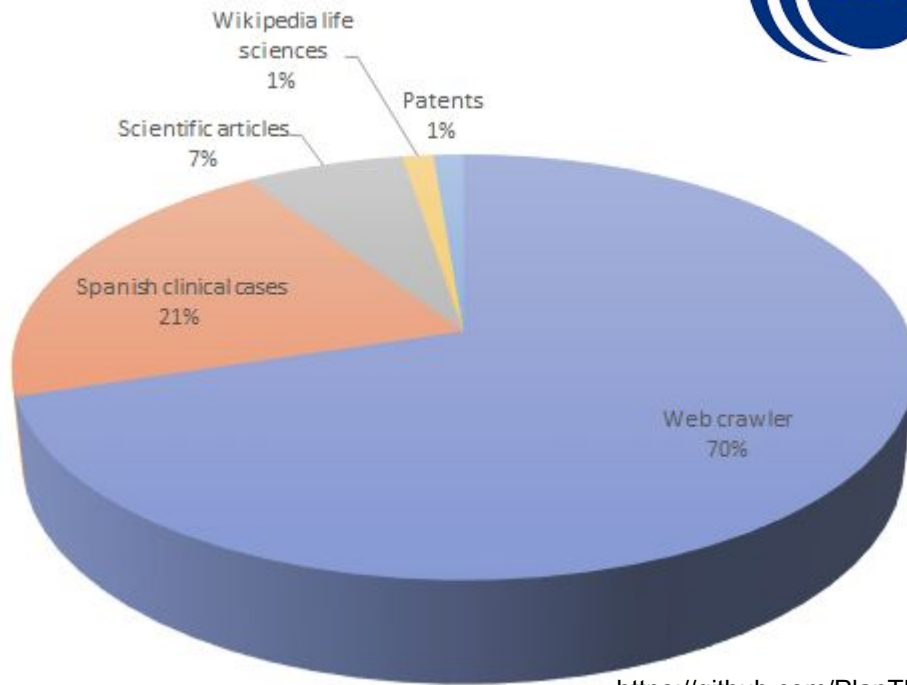
Summer School on Digital Humanities

Rodrigo del Moral, Orlando Ramos

UNAM, September 2023
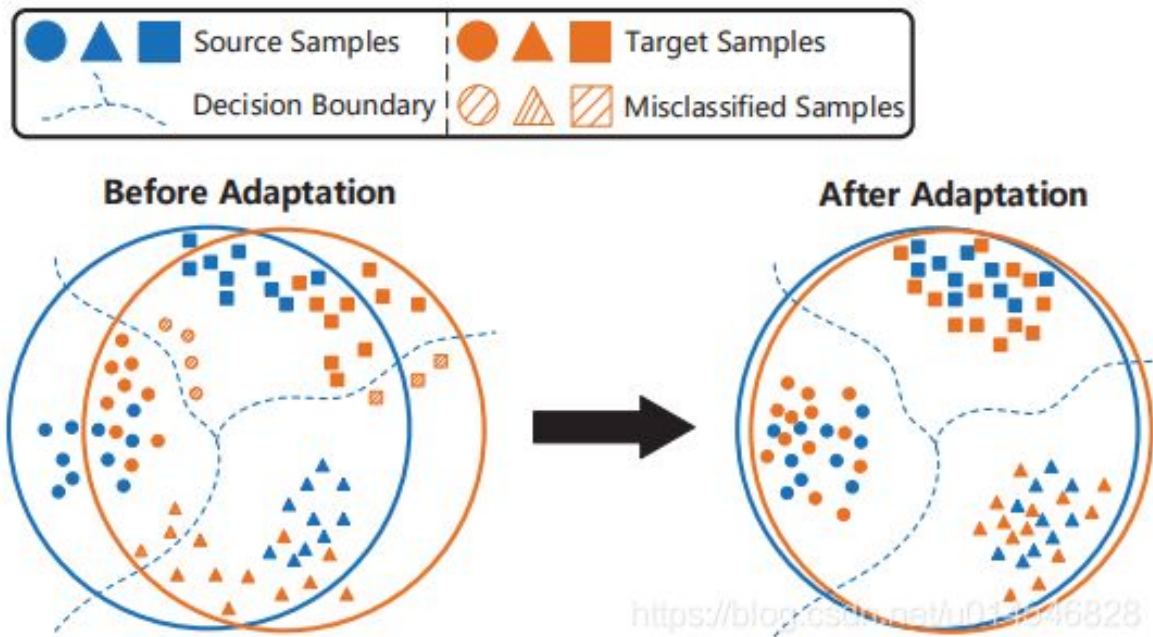
Language models for clinical text in spanish?

- roberta-biomedical-clinical-es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

https://github.com/PlanTL-GOB-ES/lm-biomedical-clinical-es

# Can we improve this LM for mexican clinical text?

- Domain adaptation

What data do we have to adapt the LM?

- Mexican clinical notes dataset
  - 78,616 notes
  - 98% Emergency services
  - 2% (General Surgery, Pediatrics, Obstetrics Gynecology, Internal Medicine, Traumatology and orthopedics, Psychology, Ophthalmology, General medicine, Otorhinolaryngology, etc)

- Chilean waiting list dataset
  - 9,000 notes
  - Only specialty consultation referral notes

Pipeline
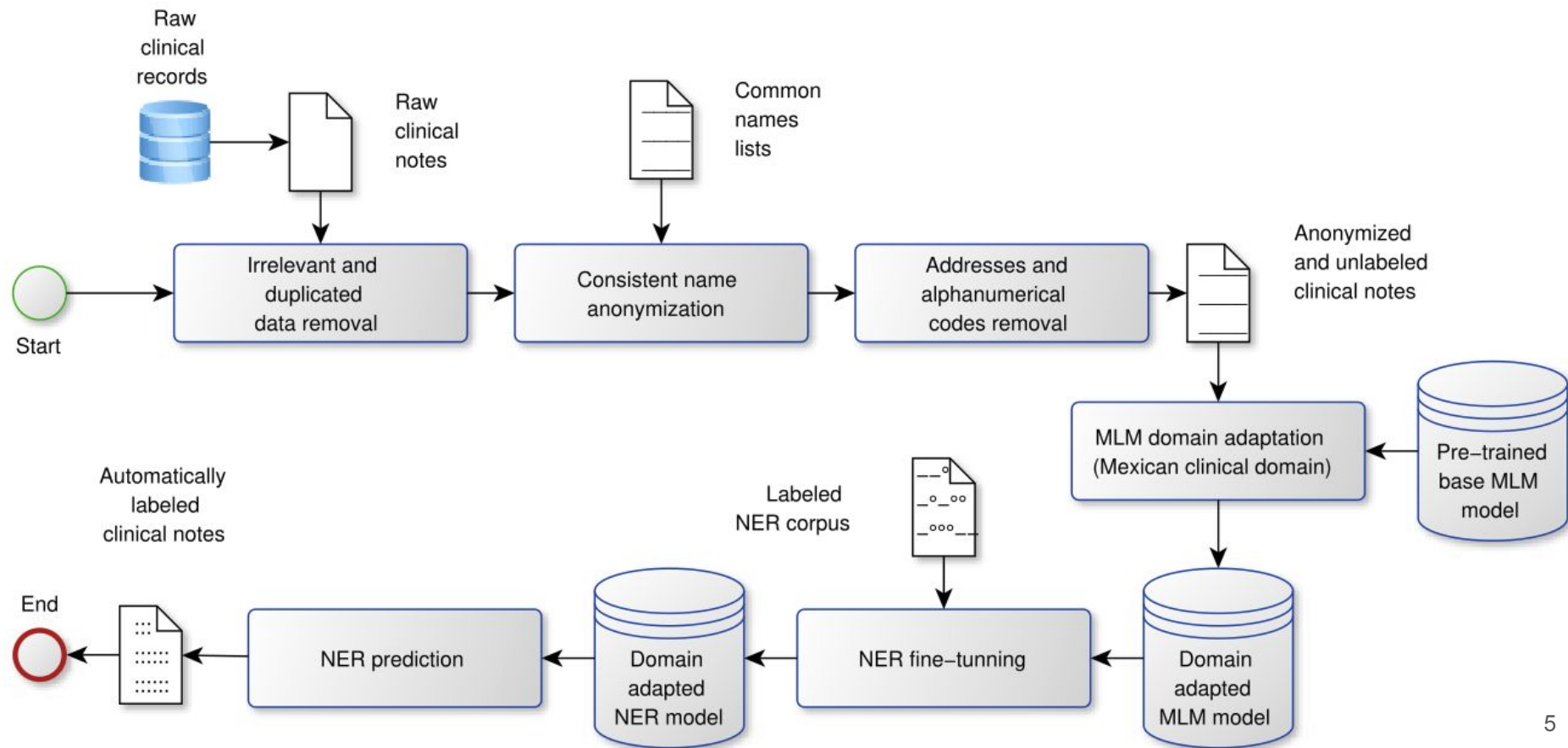
How do we evaluate this adapted LM?

- Named Entity Recognition
  - For clinical entities

- PharmaCoNER dataset
  - 1,000 Spanish clinical cases
  - Sourced from various specialties
  - Tagged with chemicals (proteins, medicines, among others)

Results

- Precision => Correctly identified entities, disregarding false negatives
- Recall => Correctly identified entities, disregarding false positives
- F1 => Good balance of Precision and Recall

PharmaCoNER			
Model	P	R	F
bsc-bio-ehr-es	**0.8921**	0.9081	0.9000
bsc-bio-ehr-es + cwl	0.8703	0.9119	0.8906
bsc-bio-ehr-es + ours	0.8826	0.9119	0.8970
bsc-bio-ehr-es + cwl + ours	0.8876	**0.9146**	**0.9009**

Tutorial: Transformers for Medical Entity Detection

Summer School on Digital Humanities

Rodrigo del Moral, Orlando Ramos

UNAM, September 2023

What about clinical NER?

"A 73-year-old woman was admitted for scheduled surgery for gastric adenocarcinoma after neoadjuvant chemotherapy, undergoing subtotal gastrectomy with reconstruction by Roux-en-Y gastrojejunostomy. As a previous history there was a probable allergy to vitamin B (Neuromade®)."

What kinds of entities can we find in clinical notes?

- Symptoms
- Diseases
- Procedures
- Lab tests / results
- Drugs
- Anatomy
- Other chemicals

We need to develop specific NER models, but how?

1	Pre-trained language model

- Trained with non-labeled data
- Preferably language and domain specific

2	Labeled NER corpus

- Annotated by experts of a specific field
- Representative of the entities that the NER model will process

3	Classification algorithm

- Good performance
- Computationally efficient
- Robust

Please open the notebook

https://shorturl.at/ezCMY