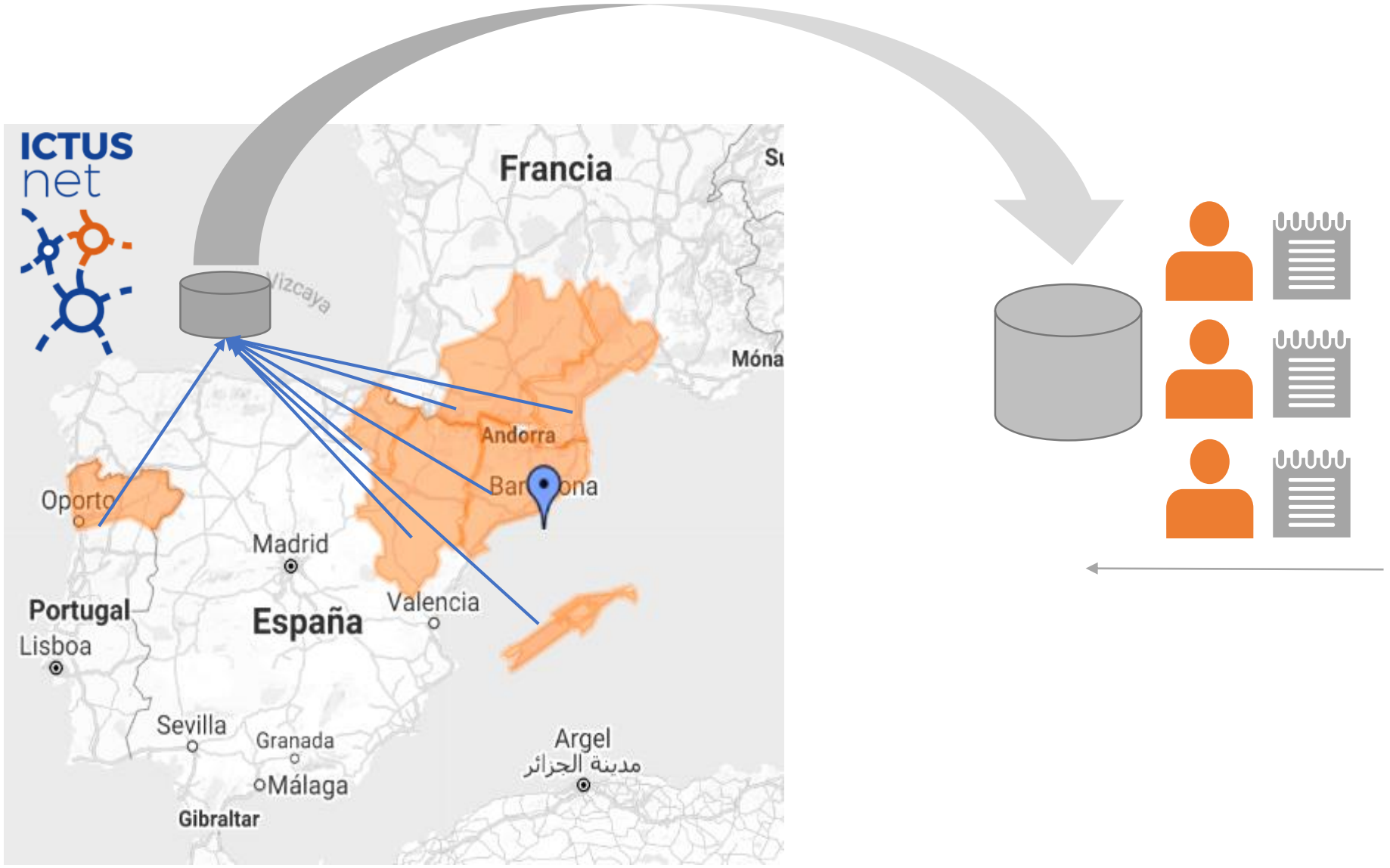
Machine learning algorithms for Natural Language Processing in stroke discharge reports

ictusnet-sudoe.eu

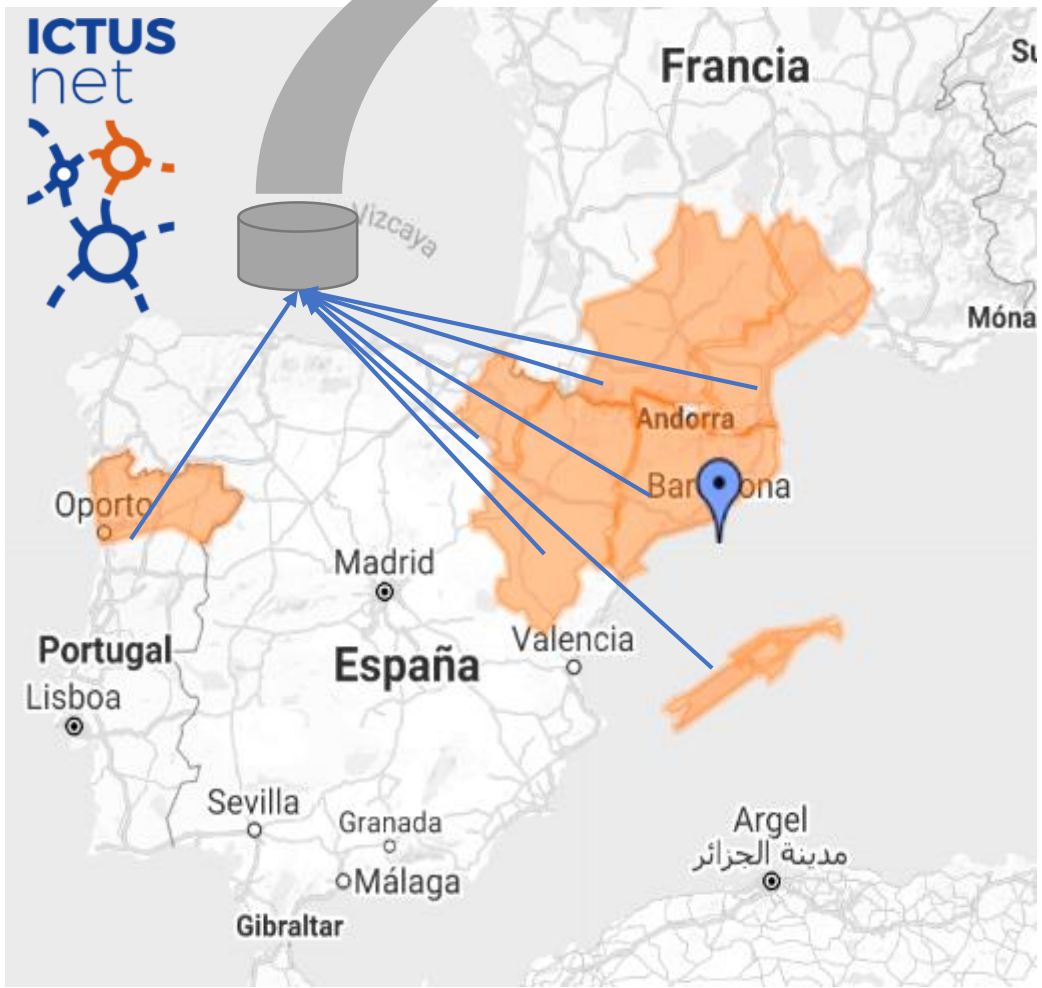
Marta Villegas
Aitor González, Siamak Barzegar,
Casimiro Carrino, Jordi Armengol, Asier Gutiérrez

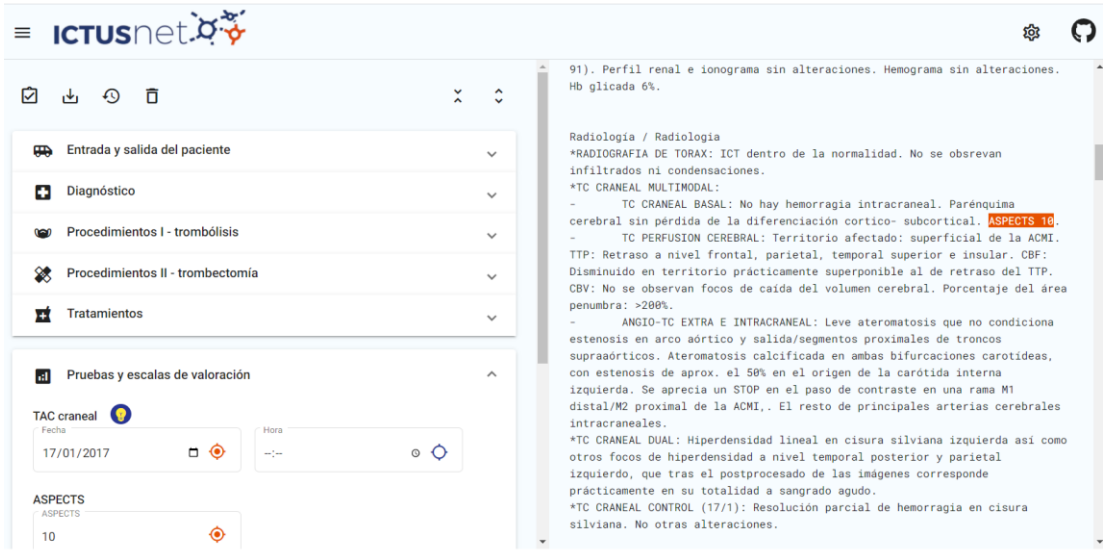
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

An information extraction service to support human experts



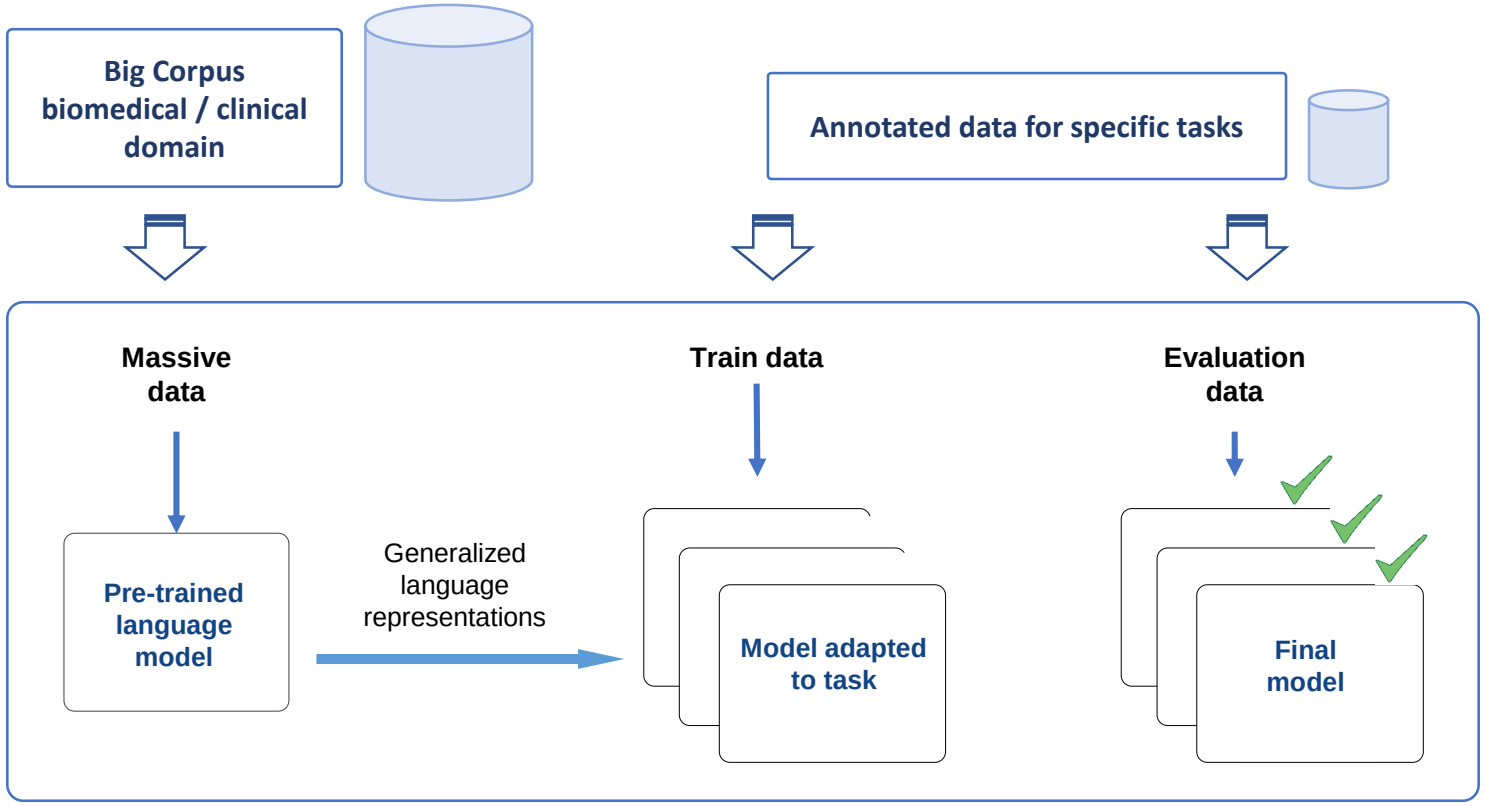
An information extraction service to support human experts





Methodology

1. We used deep learning techniques and deep neural networks to generate a **domain specific language model**.
2. Then, we adapt this model (**fine tune** it) to a specific task (Named Entity Recognition, NER)



Methodology

1 — Automatic pre-annotation

Rule base pre-annotations to assist the manual annotation process.

Software tools

2 — Manual annotation

Experts validate and/or modify the pre-annotations to generate the Gold Standard annotations used to train & evaluate models.

Gold Standard annotations

3 — Deep learning Language modelling

Large biomedical data set is used to generate a biomedical pre-trained language model. (~6GB & ~1000M tokens)

Biomedical language model

4 — Deep learning Language modelling

Clinical textual data is used to retrain de biomedical model to adapt it to the clinical setting. (120MB, almost half from ICTUSnet)

Clinical language model

5 — Deep learning Fine tuning

The final model is fine tuned to a NER task using the Gold Standard.

NER application

INDEX

1. Annotation variables & main challenges
2. (Pre)-annotation task
3. Language modelling & fine-tuning to NER
4. Evaluation
5. Summary & Conclusions
6. Examples & demo

VARIABLES and main challenges

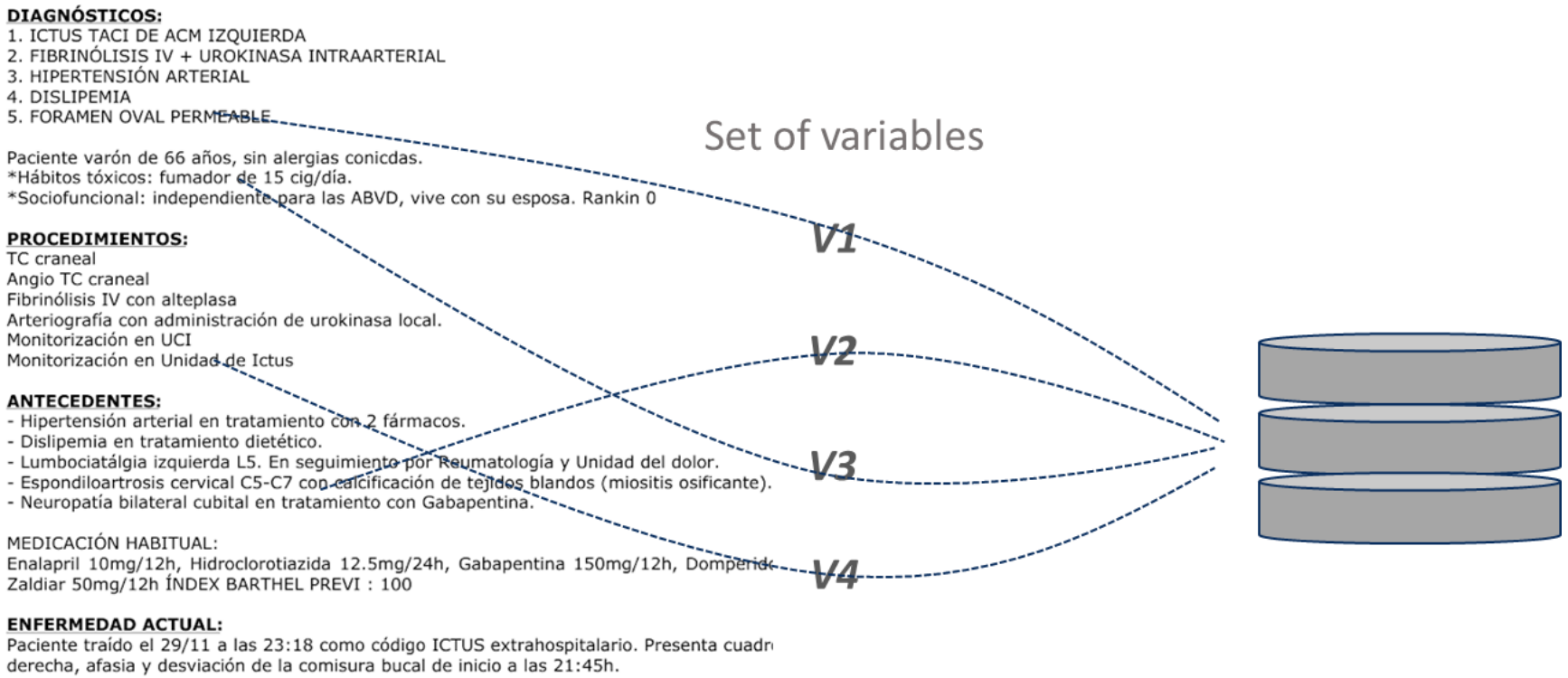
ICTUSnet includes **51 different variables of interest** to be identified in discharge clinical reports.

These can be grouped into 6 groups:

- Section headers
- Main diagnosis & associated variables
- Procedures & associated temporal information
- Treatments
- Scales
- Temporal variables

Section headers (need to normalize)

- Data coming from different providers with different structures and formats, need to normalize (use of Arquetipos suggested by Spanish Health Ministry as a mapping standard)).
- Some variables are ‘context dependent’ (they are only relevant provided they are in a specific section)
- We need *zoning* to discriminate variables by context



Section headers (and structural heterogeneiety)

SECCION_MOTIVO_DE_INGRESO
MOTIU D'INGRÉS
Paciente que ingresa por ictus ACM izq y ACA izq cardioembólicas.

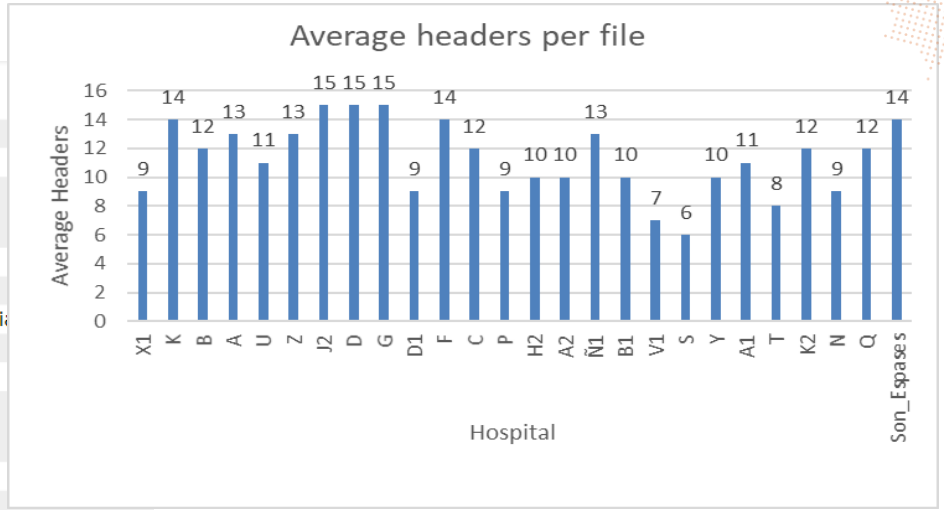
SECCION_ANTECEDENTES
ANTECEDENTS
No alergias conocidas No hábitos tóxicos FRCV:
-No HTA, no DLP, no DM
-Cardiopatías: FA anticoagulada con Sintrom (control en privada Dr. _NAME_ centro.), valvulopatía mitral no tributaria
-Niega otros antecedentes de interés.

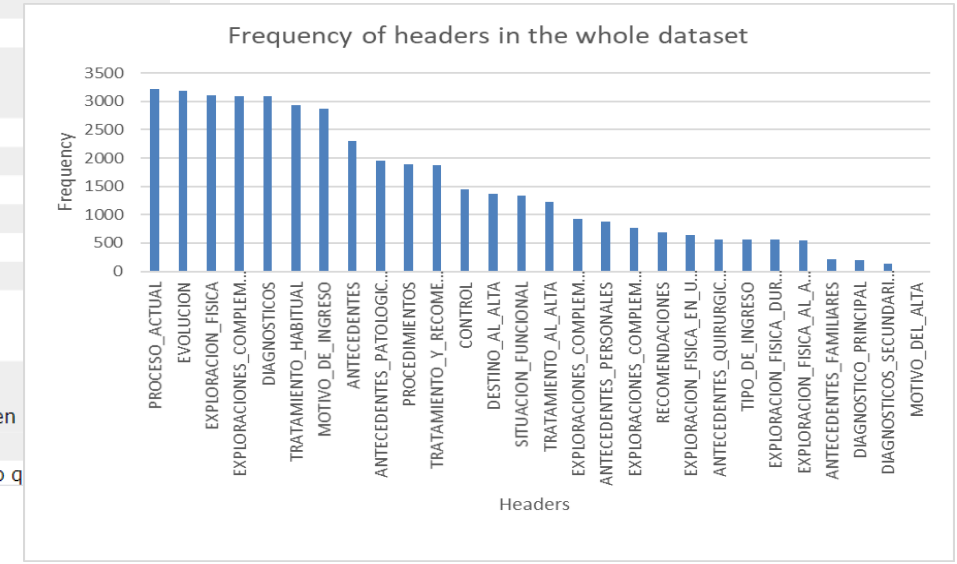
SECCION_ANTECEDENTES_QUIRURGICOS
IQ:
osteosíntesis húmero proximal izq. apendice y colecistectomía. mrankin 0.
Vive sola.
Deamulación autónoma.
Independiente para las ABVD.

SECCION_TRATAMIENTO_HABITUAL
MEDICACIÓ HABITUAL
CO-VALS FORTE 160MG/25MG VALSARTAN+DIURETIC 1 x 24 h.
Indefinida EMCONCOR COR 2,5MG BISOPROLOL, FUMARAT 1 x 24 h.
Indefinida SINTROM 4MG ACENOCUMAROL 1 x 24 h.
Indefinida

SECCION_PROCESO_ACTUAL
PROCÉS ACTUAL

Fecha_llegada_hospital
Paciente de 82 años que el día 11/01 es encontrada por familiares a las 19.40h en domicilio acostada en
quien objetiva RACE 8 y activa código ictus desde domicilio a las 20.15h.
La paciente estaba vestida con ropa de cama (ella siempre va a la cafetería a desayunar y hoy no la han visto, por lo q





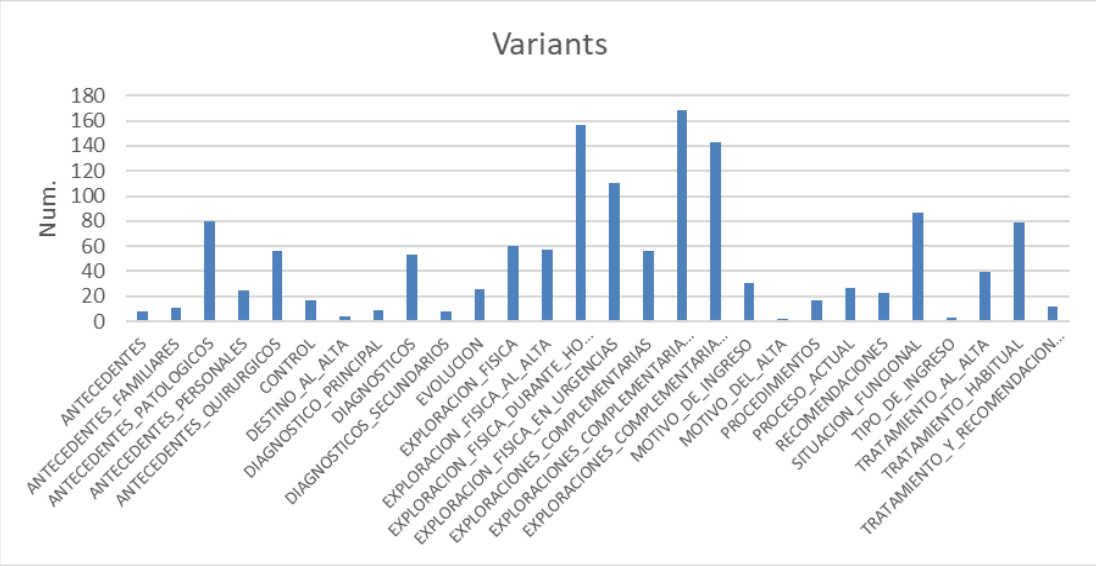
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Section headers (and naming conventions)

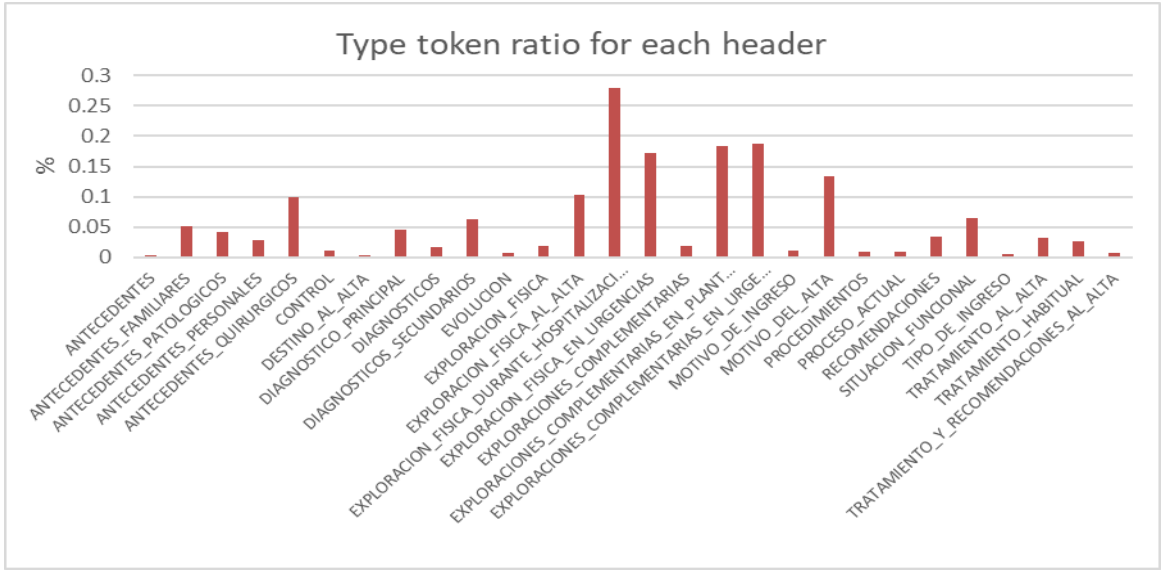
ANTECEDENTES_PATOLOGICOS				
A. patológicos		ANTECEDENTES PATOLÒGICOS		ANTECEDETES PATOLÓGICOS
Anetecedentes médicos		Antecedentes patològicos		ANTECEDETNES MÉDICOS
ANTECDENTES PATOLÓGICOS		ANTECEDENTES PATOLOGICS		Antecedetnes Patológicos
ANTECDENTES PATOLÓGICOS		ANTECEDENTES PATOLÓGICS		ANTECEDNETES PATOLÓGICOS
ANTECECENTES MÉDICOS		ANTECEDENTES PATOLÒGICS		Antecednetes Patológicos
ANTECEDENETS PATOLOGICOS		ANTECEDENTS PATOLOGICS		ANTECENDENTES PATOLÓGICOS
Antecedente Patologicos		Antecedents patològics		Antecendentes patológicos
ANTECEDENTE PATOLOGICOS		ANTECEDENTS PATOLÒGICS		ANTECENDETES PATOLÓGICOS
ANTECEDENTE PATOLÓGICOS		ANTECEDENTS MÉDICOS		ANTECENTES PATOLÓGICOS
ANTECEDENTES PATOLÓGICOS		ANTECEDENTS MÉDICS		ANTECENTS PATOLÒGICS
ANTECEDENTES PATOLÒGICOS		Antecedents mèdics		Antedentes Patológicos
Antecedentes Medicos		ANTECEDENTS PATOLÒGIC		ANTEEDENTS PATOLÒGICS
ANTECEDENTES MEDICOS		ANTECEDENTS PATOLOGIC S		APatológicos
Antecedentes médicos		ANTECEDENTS PATOLÒGIC S		HISTORIAL MÈDIC
ANTECEDENTES MÉDICOS		ANTECEDENTS PATOLOGICOS		Historial mèdic
Antecedentes Médicos		ANTECEDENTS PATOLÓGICOS		HISTORIAL MEDICO
Antecedentes mèdics		Antecedents Patológicos		HISTORIAL MÉDICO
ANTECEDENTES PAOTLÓGICOS		ANTECEDENTS PATOLÒGICOS		Historial médico
ANTECEDENTES PATOLIOGICOS		ANTECEDENTS PATOLOGICS		Malalties prèvies
Antecedentes patològices		Antecedents patológics		Otros antecedentes patológicos
ANTECEDENTES PATOLOGICOS		ANTECEDENTS PATOLÓGICS		PATOLÓGICAS
Antecedentes Patologicos		ANTECEDENTS PATOLÒGICS		PATOLOGICOS
Antecedentes patologicos		Antecedents patològics		PATOLÓGICOS
Antecedentes patológicos		Antecedents Patològics		Patológicos
ANTECEDENTES PATOLÓGICOS		ANTECEDENTS PATÒLOGICS		PATOLOGICS
Antecedentes Patológicos		Antecedents patològics FRCV		PATOLÒGICS
ANtecedentes Patológicos		Antecedents patoògics		

The 80 variants for ANTECEDENTES PATOLOGICOS

Section headers (and naming conventions)

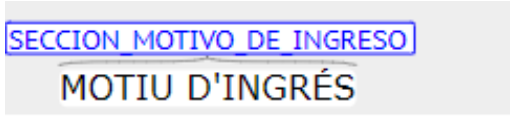


Number of variants per header (i.e. different ways to mention a header)

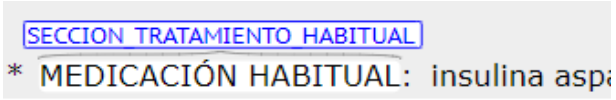


TTR is the total number of unique headers (types) divided by the total number of headers (tokens) in the documents

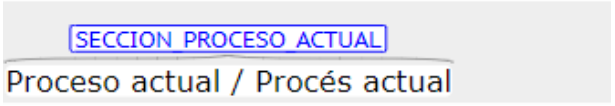
Section headers (and format heterogeniety)



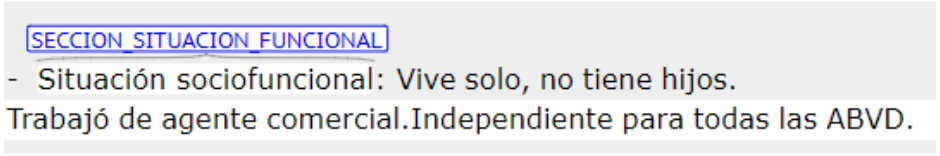
Catalan / uppercase / no special markup / full line



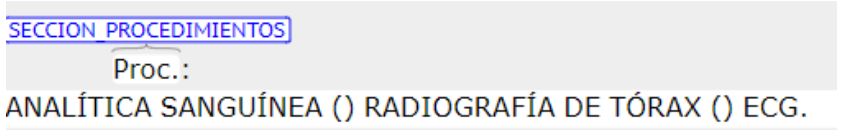
Spanish / uppercase/ special markup / inserted in paragraph



Bilingual / lowercase / no special markup / full line



Spanish / lowercase / special markup / inserted in paragraph



? / abbreviation / lowercase / no markup / full line

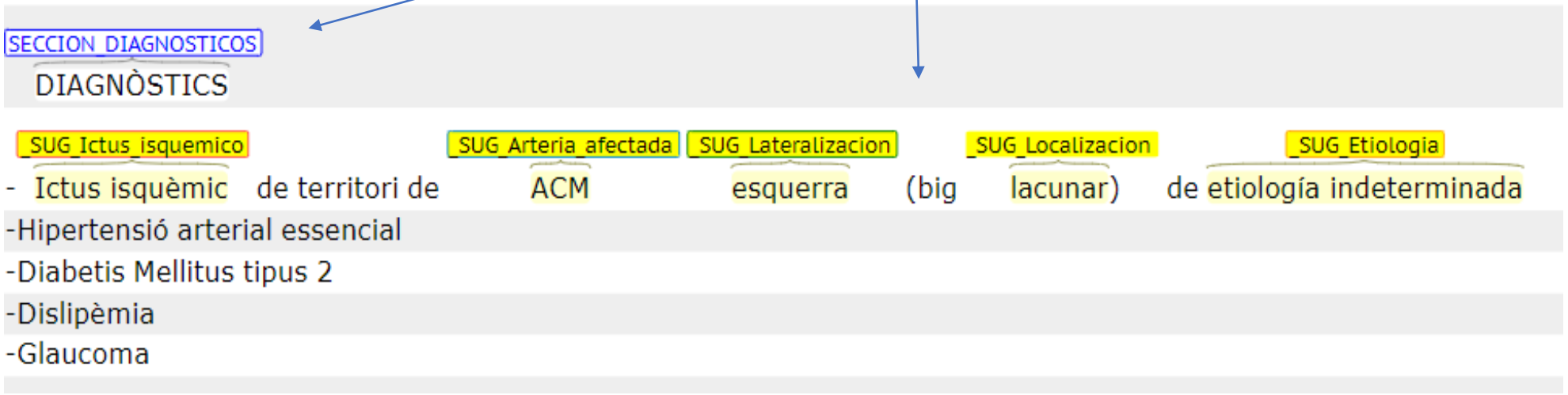
← Upper case!!

Main diagnosis and associated variables

This includes three main diagnoses*: **ictus isquémico**, **ataque isquémico transitorio** and **hemorragia cerebral*** and their associated attributes:

- affected vessel,
- localization,
- lateralization and
- etiology.

Context dependent variables
(only relevant if in DIAGNOSTICOS section)



Procedures & associated temporal information

Procedures
Trombolisis_intravenosa
Trombectomia_mecanica
Trombolisis_intraarterial
Test_de_disfagia
Tac_craneal

Associated temporal information:
Fecha / hora Tc craneal inicial
Fecha trombólisis iv
Hora inicio primer bolus de la trombólisis rtPA
Fecha trombectomía mecánica
Hora punción arterial para la trombectomía mecánica
Fecha / hora primera serie trombectomía mecánica
Hora trombólisis intraarterial
Fecha / hora recanalización
Fecha / hora finalización trombectomía
Fecha / hora trombólisis intraarterial

Hora_primer_bolus_trombolisis_rtPA
Trombolisis_intravenosa
La pacient compleix criteris per rebre tractament fibrinolític amb rtPA que s'adnministra a les 11'37h sense incidiències.

```
T30     Trombolisis_intravenosa 2568 2600       tractament fibrinolític amb rtPA
```

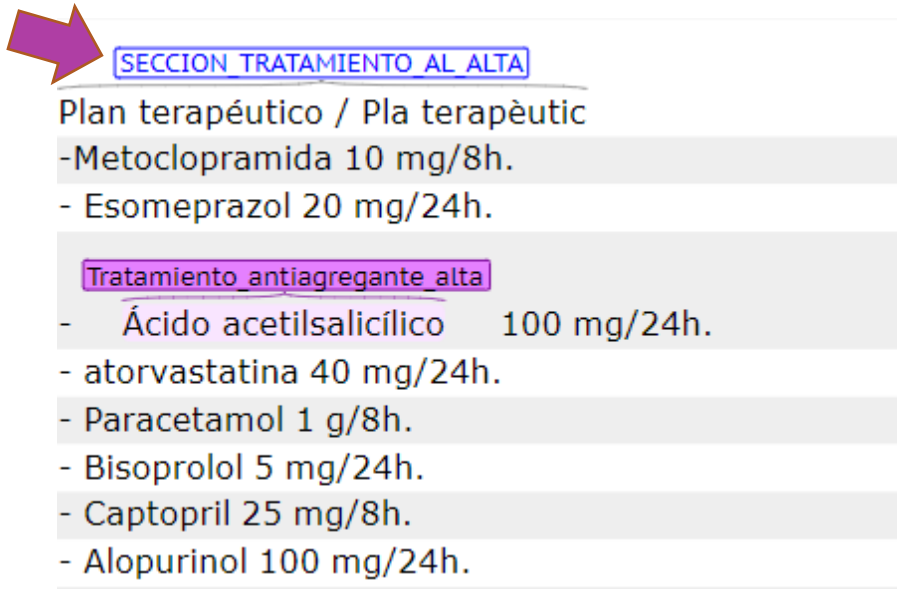
```
T14     Hora_primer_bolus_trombolisis_rtPA 2568 2631    tractament fibrinolític amb rtPA que s'adnministra a les 11'37h
#3      AnnotatorNotes T14      11:37
```

TAC_craneal Fecha_TAC TAC_craneal
-TC craneal de control a las 24h (22/3/17): En valoración comparativa con estudio TC previo no identifico

Treatments

For **treatments**, the objective is to find **anticoagulants** and **antiaggregants** and to classify them as "**pre admission medication**" or "**discharge medication**".

This task is essentially a NER task that includes a classification part (pre-admission vs discharge). This classification mostly depends on the context of the mention (i.e. the section in which the medication is listed).



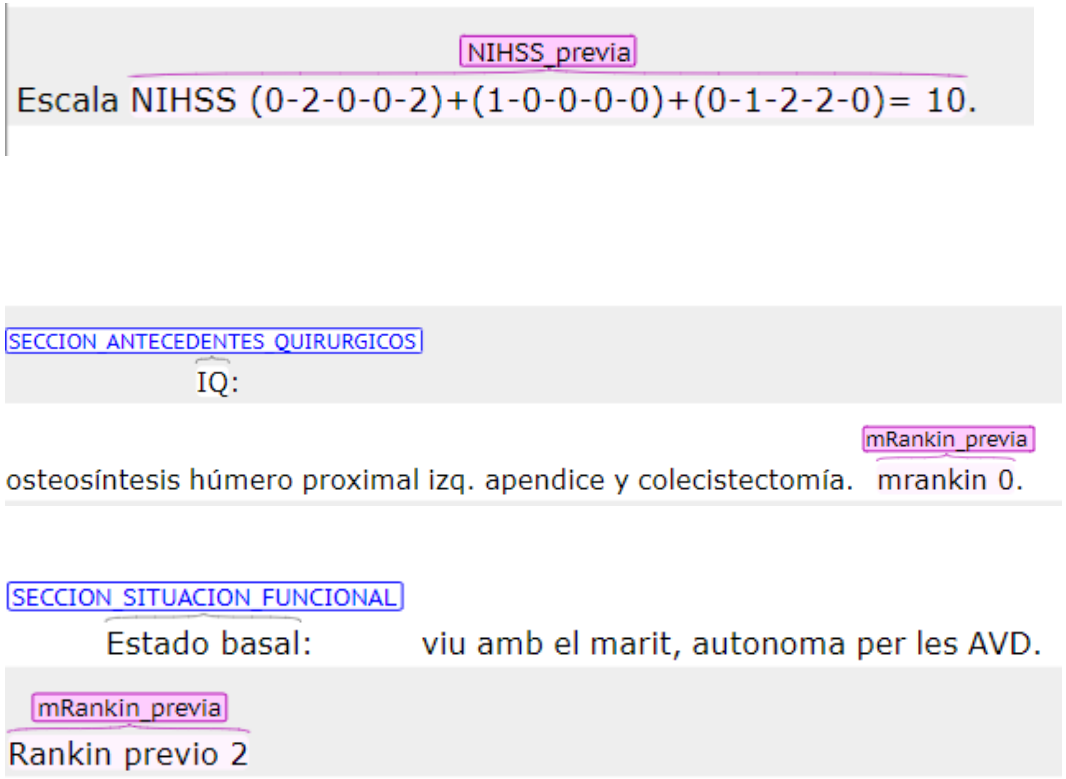
Scales

The relevant **scales** to be annotated include:

ASPECTS
mRankin_previa, mRankin_alta
NIHSS_previa, NIHSS_alta

The main challenge in this case is to find the numerical value of the rating scale (often this comes in a complex format) and to distinguish between *pre-admission vs discharge* categories.

Note that in the vast majority of cases, this distinction is not explicitly expressed in the reports. Again, these are context dependent variables.



VARIABLES and main challenges

- Multiple providers
  - Heterogeneous data with different
  - formats, structure, headers, naming conventions, …
- Nature of Language
  - Acronyms and abbreviations.
  - Ambiguity.
  - Negation
- User generated content
  - Ungrammatical/telegraphic sentences.
  - Text in different languages.
  - Text written by different personnel.
  - Typos.

Mujer de 84 años sin **ACM**. Niega hábitos tóxicos. Parcialmente dependiente **ABVD**. Vive en residencia. Antecedentes de **HTA**, **DLF** y **FA** antiagregada. Ictus **POCI ACP** izquierda en 2008, etiología cardioembólica.

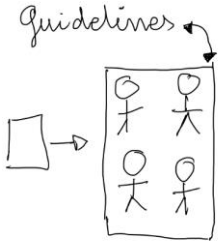
Mujer de 84 años sin **alergias medicamentosas conocidas**. Niega hábitos tóxicos. Parcialmente dependiente **adriamycin bleomycin vinblastine and dacarbazin**. Vive en residencia. Antecedentes de **hipertensión arterial**, **depresión a largo plazo** y **fibrilación auricular** antiagregada. **Ictus circulación posterior, arteria cerebral posterior** izquierda en 2008, etiología cardioembólica.

TC craneal urgente multimodal: No es visualitzen signes hemorràgics aguts intra ni extraxials.
TC de control (24h post tratamiento fibrinolítico): Infart fronto-parieto-temporal dret establert, en territori d'ACM dreta, sense signes de transformació hemorràgica
Fàcil masticació. 1º Plato sopa o pure.
Durante el ingreso en la unidad de ictus, presenta un empeoramiento neurológico, por lo que se realiza un TC craneal de control que evidencia una nueva lesión isquémica en ACP izquierda. También presenta una fibrilación auricular de debut, por lo que se decide iniciar tratamiento anticoagulante. Avui control INR (s'enviarà resultat a IAS).
AngioTC craneal (6.12.2017): Poligono de Willis normoconfigurado, sin imagenes de adiccion ni defectos de replección. No se observan imagenes sugestivas de malformación vascular. Segmento V4 esquerre hipoplàsico.

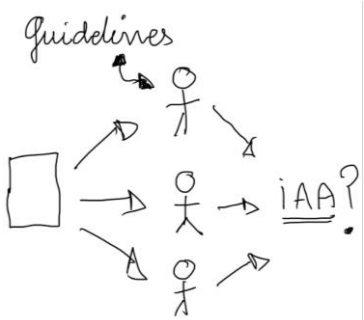
Annotation task

Annotation task is difficult and needs good annotation guidelines, good protocols and a friendly annotation tool.

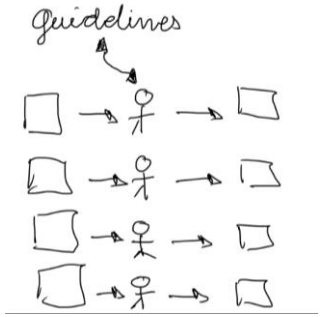
1

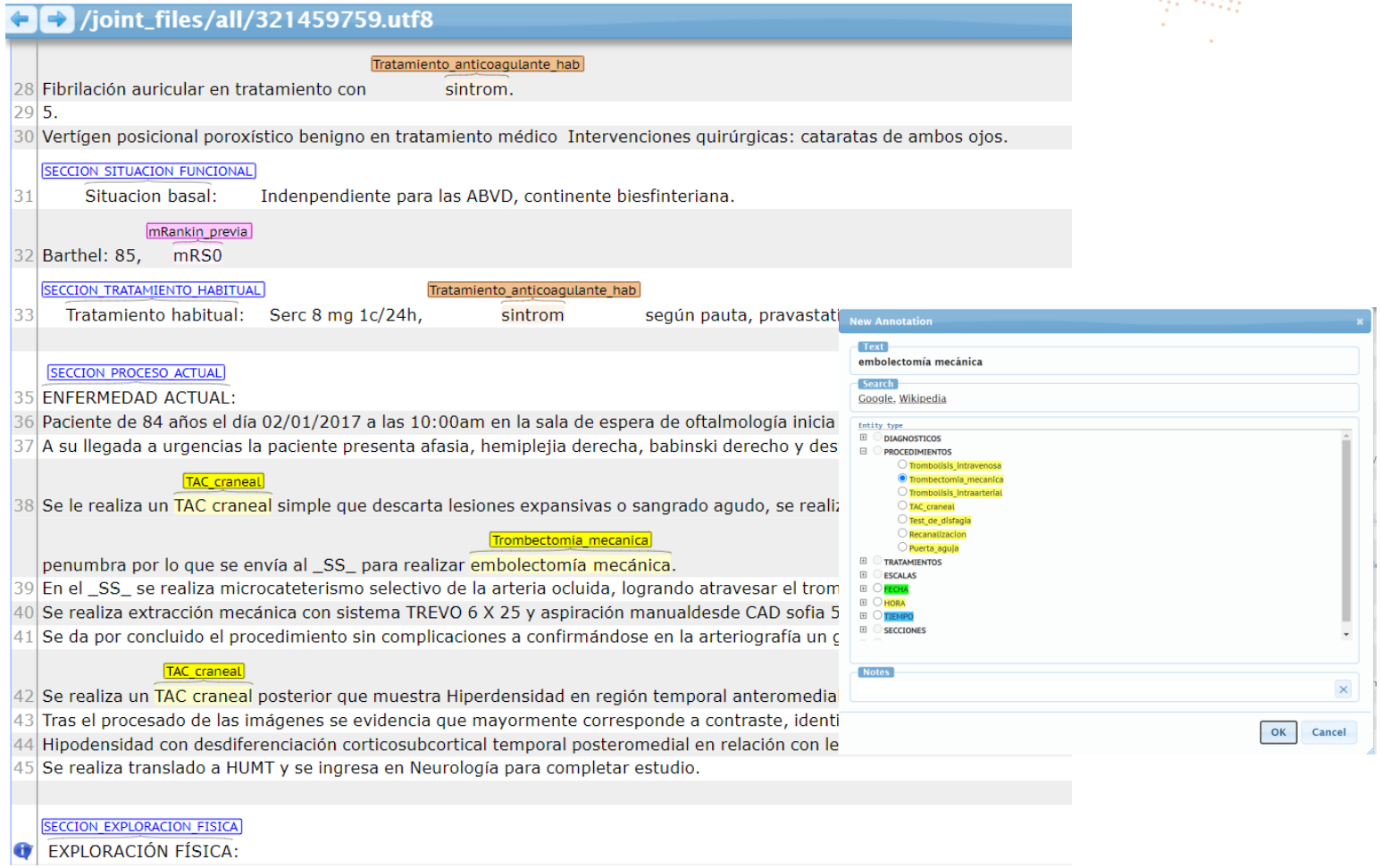


2



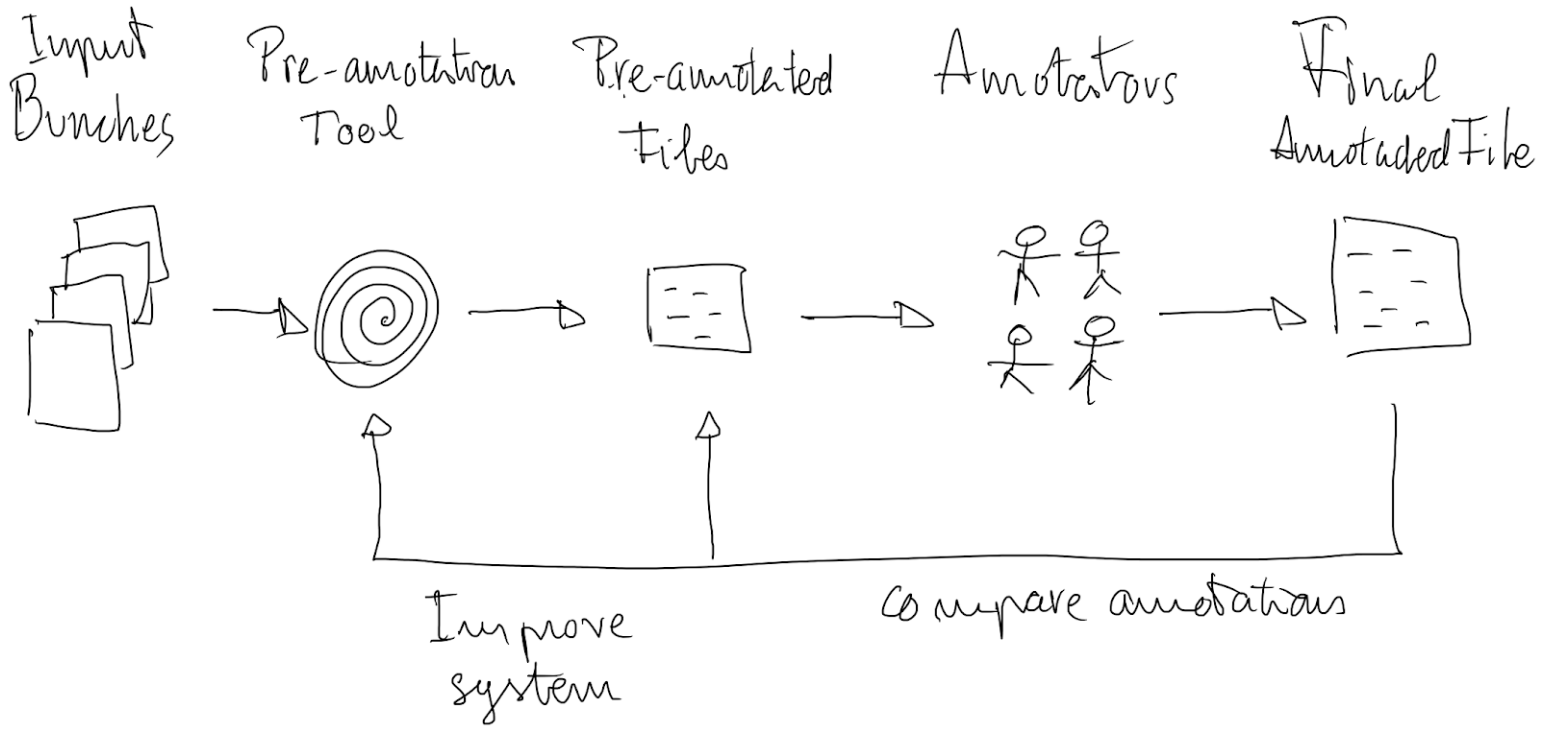
3





(Pre-)annotation

To ease the task, an **automatic pre-annotation system** was developed in an **iterative way**, so that the process was split into different steps, each consisting of ~100 discharge reports. At each new bunch of annotated files, the system is evaluated against the human annotations and modified to improve its performance



(Pre-)annotation (evaluation)

Num examples	True positives	False positives	False negatives	Accuracy	Precision	Recall	F1
5710	5012	894	698	0.759	0.849	0.878	0.863

Global average results using the Test set

Language Modelling

3 We generated a **RoBERTa-base model** with 12 layers/heads and 768 hidden layer sizes for a total number of 126M parameters.

- We kept the original Roberta hyperparameter configuration and trained with a **masked language model** objective.
- The model was trained for 48 hours using 16 NVIDIA V100 GPUs of 16GB DDRAM.
- After training, we selected as the best model the checkpoint that achieved the lowest perplexity.

4 Then, we adapted the model to the clinical domain by **overtraining it** with 120MG of clinical textual data (including nearly 34MB of ICTUSnet data provided by AQuAS, Son Espases and IACS).

- We continued the training process for 48h more using the learning rate value reached by the best checkpoint trained on the biomedical corpora.

5 Finally, we **fine-tuned** our pre-trained models for NER task using the ICTUSnet Gold Standard dataset.

- The gold standard was split into train, dev and test sets with standard proportions: 80% for training (656 documents), 10% for valid (83 documents), 10% for test (83 documents).
- We fine-tuned for 10 epochs and selected the best epoch validating on the dev set.

3

Deep learning
Language modelling

Large biomedical data set is used to generate a biomedical pre-trained language model.
(~6GB & ~1000M tokens)

Biomedical language model

4

Deep learning
Language modelling

Clinical textual data is used to retrain de biomedical model to adapt it to the clinical setting.
(120MB, almost half from ICTUSnet)

Clinical language model

5

Deep learning
Fine tuning

The final model is fine tuned to a NER task using the Gold Standard.

NER application

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Language Modelling

We collected a big **biomedical corpora** gathering from a variety of medical resources, namely scientific literature, clinical cases and crawled data.

We cleaned each corpus independently using a **cleaning pipeline** with customized operations designed to read data in different formats, split into sentences, detect the language, remove noisy and bad-formed sentences, …

Finally we **deduplicate** and eventually output the data with their original document boundaries.

Corpus name	Text Size (GB)	Final size (GB)	Raw tokens	Cleaned tokens	Num. sentences
Clinical cases cardiology	0.0035	0,001	149,904.00	147,790.00	9,970.00
Clinical cases radiology	0.0066	0,001	177,366.00	170,997.00	9,948.00
libros_casos_clinicos	0.0083	0,007	1,137,555.00	1,024,797.00	68,833.00
Clinical cases COVID	0.0084	0,001	82,201.00	82,091.00	3,896.00
EMEA corpus	0.087	0,034	13,797,362.00	5,377,448.00	284,575.00
Patents	0.087	0,084	14,022,520.00	13,463,387.00	253,924.00
wikipedia_life_sciences	0.172	0,088	18,771,176.00	13,890,501.00	832,027.00
barr2_background	0.188	0,159	28,868,022.00	24,516,442.00	1,029,600.00
Pubmed	0.211	0,013	1,957,479.00	1,858,966.00	103,674.00
REEC (casos clínicos)	0.823	0,028	4,581,755.00	4,283,453.00	220,726.00
mespen_medline	1.2	0,38	6,864,901.00	4,166,077.00	322,619.00
pdfs_general	3.3		09,124,996.00	7,146,139.00	5,252,481.00
Scielo	3.891	0,631	61,837,972.00	60,007,289.00	2,668,231.00
Medical crawler	606	4,5	?	746,368,185.00	32,766,976.00
TOTAL	**615.9858**	**5,927**	**261,373,209.00**	**972,503,562.00**	**43,827,480.00**

Language Modelling

We used the resulting corpus to train a **RoBERTa-base model** with 12 layers/heads and 768 hidden layer sizes for a total number of 126M parameters.

- We kept the original Roberta hyperparameter configuration and trained with a **masked language model** objective.
- The model was trained for 48 hours using 16 NVIDIA V100 GPUs of 16GB DDRAM.
- After training, we selected as the best model the checkpoint that achieved the lowest perplexity.

Then, we adapted the model to the clinical domain by **overtraining it** with 120MG of clinical textual data (including nearly 1GB of ICTUSnet data provided by AQuAS, Son Espases and IACS).

- We continued the training process for 48h more.

Finally, we **fine-tuned** our pre-trained models for NER task using the ICTUSnet Gold Standard dataset.

- The gold standard was split into train, dev and test sets with standard proportions: 80% for training (656 documents), 10% for valid (83 documents), 10% for test (83 documents).
- We fine-tuned for 10 epochs and selected the best epoch validating on the dev set.

We used both, the Biomedical model and the Cinical model to generate and compare the predictions

Evaluation

Annotations are converted from **BRAT** standoff format to **BIO/IOB** (beginning, inside, outside) format.

- the prefix "B" in front of a Tag indicates the beginning of a chunk,
- an "I" indicates that we are still inside that chunk.
- the "O" tag is used to indicate that a token does not correspond to any of the entities to be tagged.

Given the high number of tokens with to the class "O", we do not consider them when predictions and GS we have an O label

this avoids raising the result due to the fact that O is the majority class, i.e., that the vast majority of tokens do not belong to any of the entities.

In this example, only the grey raws are evaluated. (in red the wrong ones and in Green the good ones).

Once the lines with double O are removed, we evaluate the model using **accuracy**, **precision**, **recall** and **F1**

Accuracy: number of correct predictions / total number of predictions
Precision: true positives / predicted positives (how many selected items are relevant)
Recall: true positives / actual positives (how many relevant items are selected)
F1: (balance between precision and recall)

token	tag
Vive	O
con	O
su	O
esposa	O
,	O
independiente	O
para	O
ABVD	B-Antecedente
,	O
mRs	B-mRankin_previa
0	I-mRankin_previa
.	O

Evaluation

Annotations are converted from **BRAT** standoff format to **BIO/IOB** (beginning, inside, outside) format.

- the prefix "B" in front of a Tag indicates the beginning of a chunk,
- an "I" indicates that we are still inside that chunk.
- the "O" tag is used to indicate that a token does not correspond to any of the entities to be tagged.

Given the high number of tokens with to the class "O", we do not consider them when predictions and GS we have an O label

this avoids raising the result due to the fact that O is the majority class, i.e., that the vast majority of tokens do not belong to any of the entities.

In this example, only the grey raws are evaluated. (in red the wrong ones and in Green the good ones).

Once the lines with double O are removed, we evaluate the model using **accuracy**, **precision**, **recall** and **F1**

Accuracy: number of correct predictions / total number of predictions
Precision: true positives / predicted positives (how many selected items are relevant)
Recall: true positives / actual positives (how many relevant items are selected)
F1: (balance between precision and recall)

tken	GS	prediction
Vive	O	O
con	O	O
su	O	O
esposa	O	O
,	O	O
independiente	O	O
para	O	O
ABVD	O	B-Antecedente
,	O	O
mRs	B-mRankin_previa	B-mRankin_previa
0	I-mRankin_previa	I-mRankin_previa
.	O	O

Evaluation

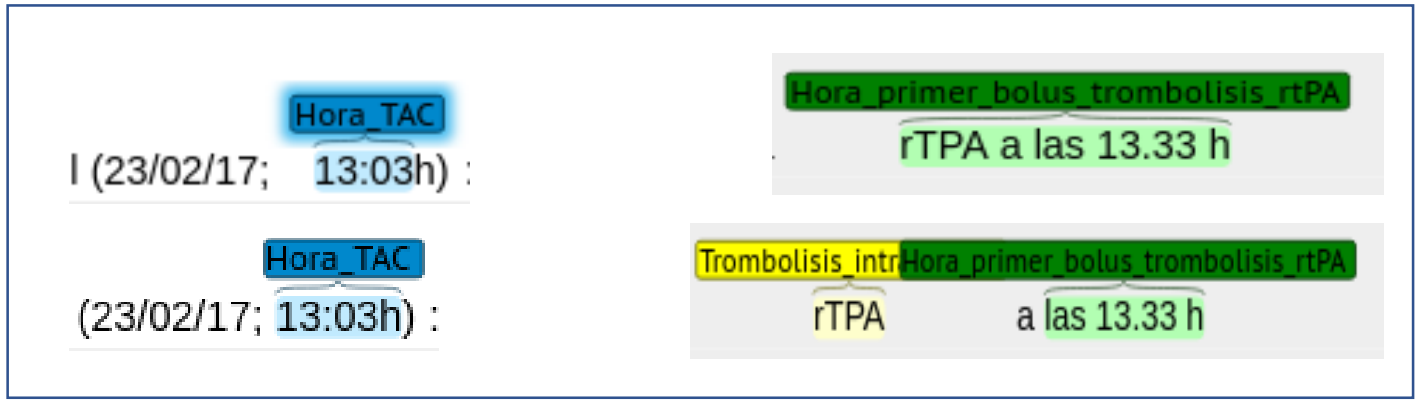
For time **variables** the evaluation is different. In this case, we have three components:

1. The textual evidence in text containing a time variable (text span)
2. The tag assigned (label)
3. The normalized time expression

When evaluating, we require that he predictions and the Gold standard

- for (2) and (3) above are equal, and
- there is some overlapping in (1) (textual evidences are long and 'irrelevant')

Examples:

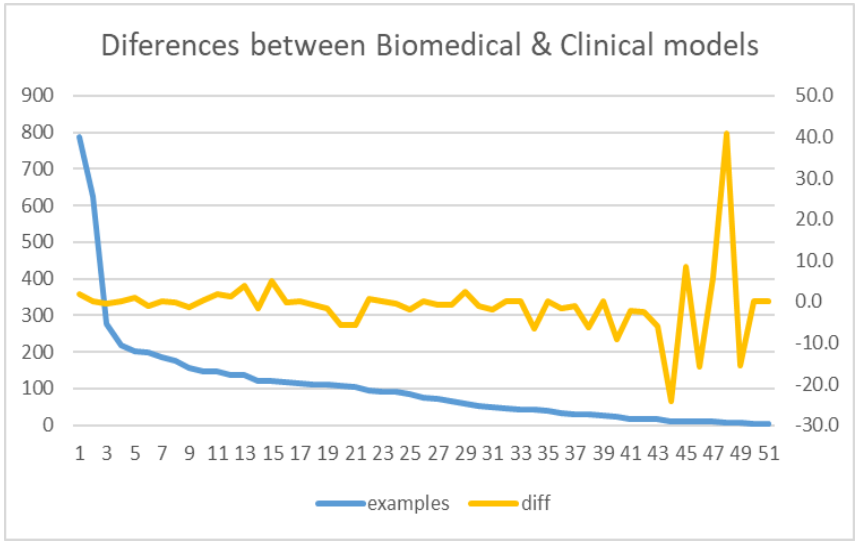


Same tags, same normalized times, overlapping mentions (note the missing *h*), **OK**

Results

Model	Examples	Tp	Fp	Fn	Acc	Pre	Rec	F1
Biomedical	5455	5125	675	330	0.836	0.884	0.940	**0.911**
Clinical	5455	5104	638	351	0.838	0.889	0.936	**0.912**

Global average results comparing Biomedical and Clinical models



Differences between Biomedical and Clinical models.

Results

tag	examples	Biomedical	Clinical	diff
NIHSS	786	0.847	0.830	1.7
TAC_craneal	625	0.983	0.983	0.0
mRankin	275	0.961	0.966	-0.5
SECCION_EXPLORACIONES_COMPLEMENTARIAS	217	0.954	0.954	0.0
SECCION_MOTIVO_DE_INGRESO	203	0.983	0.974	0.9
SECCION_TRATAMIENTO_Y_RECOMENDACIONES_AL_ALTA	200	0.980	0.990	-1.0
SECCION_PROCESO_ACTUAL	185	0.981	0.981	0.0
SECCION_EXPLORACION_FISICA	177	0.969	0.972	-0.3
SECCION_TRATAMIENTO_HABITUAL	155	0.937	0.950	-1.3
SECCION_EVOLUCION	147	0.964	0.961	0.3
Trombolisis_intravenosa	146	0.925	0.906	1.9
SECCION_ANTECEDENTES	137	0.974	0.962	1.2
SECCION_EXPLORACION_FISICA_DURANTE_HOSPITALIZACION	136	0.842	0.803	3.9
Trombectomia_mecanica	122	0.799	0.816	-1.7
SECCION_EXPLORACION_FISICA_EN_URGENCIAS	121	0.886	0.836	5.0
Ictus_isquemico	117	0.895	0.896	-0.1
SECCION_DESTINO_AL_ALTA	113	0.968	0.968	0.0
Etiologia	110	0.854	0.861	-0.7
SECCION_DIAGNOSTICOS	110	0.943	0.960	-1.7
ASPECTS	107	0.811	0.869	-5.8
SECCION_EXPLORACIONES_COMPLEMENTARIAS_EN_PLANTA_DE_NEUROLOGIA	104	0.755	0.813	-5.8
Tratamiento_antiagregante	93	0.882	0.875	0.7
SECCION_ANTECEDENTES_PATOLOGICOS	91	0.937	0.937	0.0
SECCION_TRATAMIENTO_AL_ALTA	91	0.941	0.945	-0.4
Tratamiento_anticoagulante	86	0.667	0.687	-2.0
SECCION_EXPLORACIONES_COMPLEMENTARIAS_EN_URGENCIAS	75	0.938	0.938	0.0
SECCION_SITUACION_FUNCIONAL	72	0.966	0.973	-0.7
Arteria_afectada	64	0.756	0.764	-0.8
Lateralizacion	59	0.875	0.850	2.5
SECCION_TIPO_DE_INGRESO	53	0.981	0.991	-1.0
SECCION_PROCEDIMIENTOS	50	0.961	0.980	-1.9
SECCION_EXPLORACION_FISICA_AL_ALTA	46	0.893	0.893	0.0
SECCION_ANTECEDENTES_PERSONALES	42	0.988	0.988	0.0
SECCION_RECOMENDACIONES	42	0.864	0.930	-6.6
SECCION_MOTIVO_DEL_ALTA	40	1.000	1.000	0.0
SECCION_ANTECEDENTES_QUIRURGICOS	32	0.853	0.870	-1.7
Localizacion	31	0.703	0.714	-1.1
SECCION_CONTROL	31	0.921	0.984	-6.3
Test_de_disfagia	27	1.000	1.000	0.0
Hora_TAC	22	0.750	0.842	-9.2
Tiempo_puerta_aguja	18	0.955	0.978	-2.3
Hora_primer_bolus_trombolisis_rtPA	18	0.895	0.919	-2.4
Hemorragia_cerebral	17	0.684	0.743	-5.9
Hora_recanalizacion	10	0.667	0.909	-24.2
SECCION_ANTECEDENTES_FAMILIARES	10	0.909	0.824	8.5
SECCION_DIAGNOSTICO_PRINCIPAL	9	0.842	1.000	-15.8
Hora_inicio_trombectomia	9	1.000	0.941	5.9
SECCION_DIAGNOSTICOS_SECUNDARIOS	8	0.941	0.533	40.8
Ataque_isquemico_transitorio	6	0.615	0.769	-15.4
Hora_primera_serie_trombectomia	5	0.600	0.600	0.0
Hora_fin_trombectomia	5	1.000	1.000	0.0
ALL	**5455**	**0.911**	**0.912**	**-0.097**

Conclussions

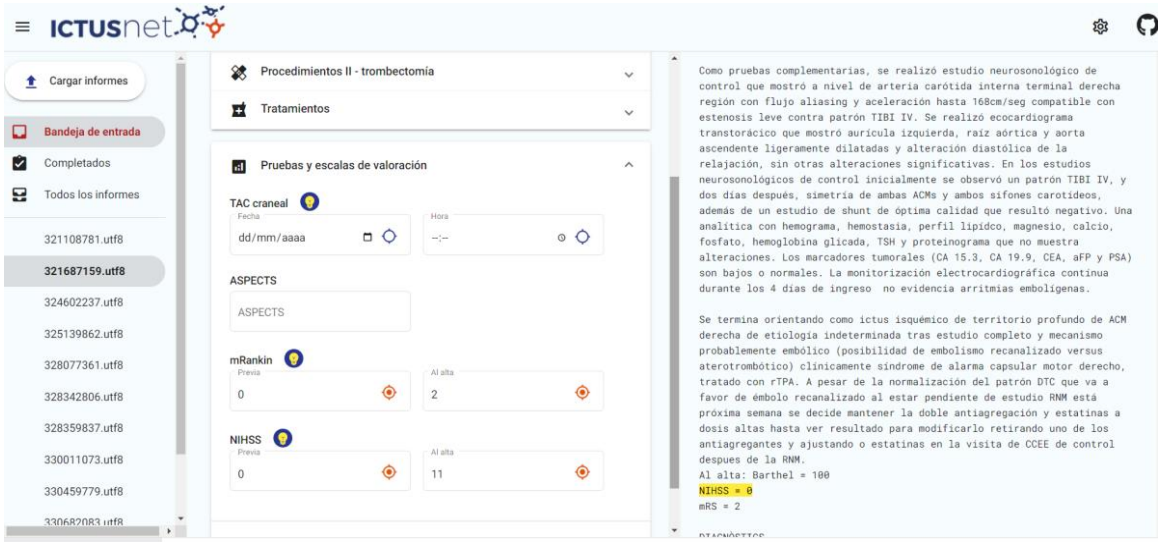
- The information extraction task was complex and ambitious, with 51 different types of variables.
- In most cases, the variables are '**context-dependen**t', which adds an extra difficulty of the task.
- **Temporal variables** are a case apart: in most cases the textual evidence shows an enormous variety. Such is the variety that, for the pre-annotation tool, we decided not to address the coding of these variables and limited ourselves to coding dates and times without going any further.
- The model managed to learn complex aspects such as the **'context sensitivity**' (this is very clear in the diagnostic variables, for example).
- The model managed to successfully learn the **complex temporal variables** that we had given up in the rule based system.

Time variable	Examples	Biomedical	Clinical
Hora_TAC	22	0.750	0.842
Tiempo_puerta_aguja	18	0.955	0.978
Hora_primer_bolus_trombolisis_rtPA	18	0.895	0.919
Hora_recanalizacion	10	0.667	0.909
Hora_inicio_trombectomia	9	1.000	0.941
Hora_primera_serie_trombectomia	5	0.600	0.600
Hora_fin_trombectomia	5	1.000	1.000

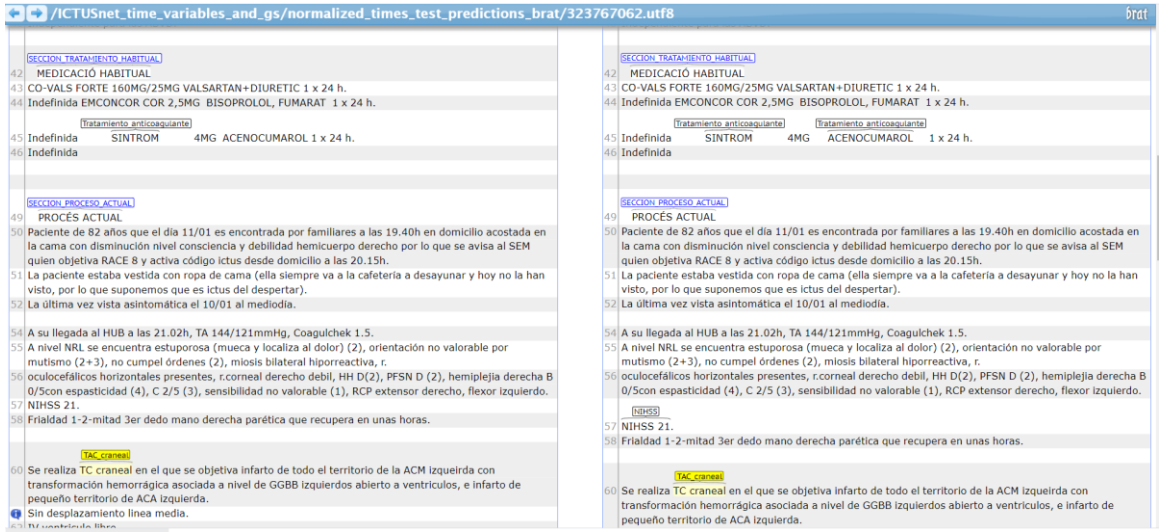
Conclussions

- We did not demonstrate that retraining with clinical data improves the model. We believe that
  - (i) more clinical data (from the stroke domain) could have improved the system,
  - (ii) mixing data from the very beginning would have positive effect, we are working on this
- **The results of the deep learning models are pretty good, reaching 91% F1 on average**. That is without applying any other (post)-process for system improvement. In this exercise we just wanted to evaluate the performance of deep learning techniques.
- **The results obtained demonstrate that the use of language technologies can be of great help in clinical information extraction tasks, as in the case of ICTUSnet.**

Demo & prototype



http://temu.bsc.es:81/



https://temu.bsc.es/ICTUSnet/diff.xhtml?diff=%2FICTUSnet_time_variables_and_gs%2Ftest_brat_gs%2F#/ICTUSnet_time_variables_and_gs/normalized_times_test_predictions_brat/323767062.utf8

#ICTUSnetBCN2020

THANK YOU

Project funded by the Interreg Sudoe Programme through the European Regional Development Fund