

Adverse Drug Reaction Detection

The Data Science at the service of Pharmacovigilance

EDSON FLOREZ, PHD STUDENT

FREDERIC PRECIOSO, MICHEL RIVEILL (I3S)

DAVID DARMON, PASCAL STACCINI (CHU DE NICE)

Adverse Drug Reaction (ADR) detection

Problem of detect correlations in the medical research domain, to establish if an adverse clinical state have appeared after taking medicines.

Input: Clinical notes contain medication, disorders, diseases, etc.

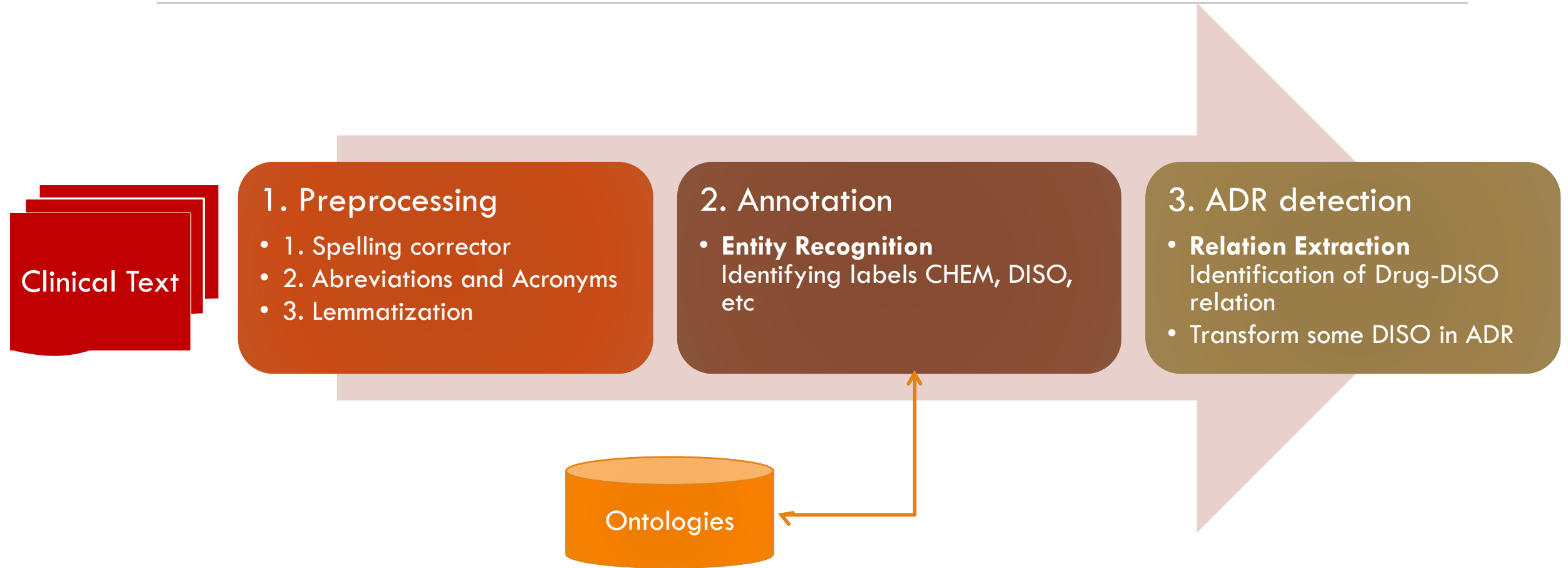
Output: Adverse Drug Reaction (ADR) of treatments

“the patient has **internal bleeding** secondary to **warfarin**”

Disorder[DISO]>>> ADR >>> **Medication[CHEM]**

ADR is an adverse event caused by a drug (Relation)

WORKFLOW



PRIMEGE Database

Collect anonymous data directly from the consultation software (EHR), with no effort of doctors to feed the database for research purpose.

Data of 13 GPs (general practitioners) collected from 2012

Data contain both structured text with codes and notes in free text (unstructured)

Volume de données

Elément	Nombre
Patients	38 970
Consultations	241 472
Antécédents	146 333
Données de biométrie	218 371
Motifs de consultation	157 009
Diagnostics	109 181
Lignes de prescriptions médicamenteuses	560 536
Symptômes recueillis	11 638
Procédures de soin	8 146
Examen complémentaires	546 262
Prescriptions paramédicales	11 048
Observations/notes	36 702

Data until 2016

- medical observation and symptoms
- weight, blood pressure
- diagnoses
- reasons of encounter
- radiology results
- data from another doctor
- **Adverse Drug Event**

1. Pre-processing

Denoising of PRIMEGE notes

- Abbreviations and medical jargon
 - cardio instead: cardiologue cardiologique cardiomégalie cardiopathie
 - gastro instead: gastroenterite gastrolenterolgue gastroscopie gastrocnemiens
 - pulm instead pulmonaire, cardiopulm, cardiopulmonaire
 - nl=normal, g. =gauche, qq=quelques, dte=droite, trt=traitement, tr=trouble
- Correction of wrong spelling words
 - apetit instead appétit
 - esport instead sport
- Medical Acronyms
 - TA, AB, TDR, ADP, ASD, MT, RC, EFR, etc

2. Annotation

Extracting Concepts with Multiple Terminologies (**ECMT**) tool** by **CISMef*** (Rouen CHU), which annotates text using the concepts of health ontologies/terminologies in French.

Cholestases intrahépatiques fibrogènes familiales et anomalies héréditaires du métabolisme hépatocyttaire des acides biliaires

Effacer 1 phrases annotées en 192 ms. 4 codes distincts identifiés.

Codes identifiés

Terme	Ter.	Code	CUI	Cond.	Cbxt.
acides et sels biliaires	MSH	D001647	C0005391		
cholestase intrahépatique	MSH	D002780	C0008372		
E70-E90 anomalies du métabolisme	ICD	E70-E90			
héréditaire	NCI	C27998	C0439660		

- Terme: preferred Term (for synonyms)
- Ter (Terminology acronym): 55 terminologies included in the repository **HeTOP** [11]
- Code: Internal code of the terminology
- CUI (Concept UMLS Identifier): Unified Medical Language System (UMLS) code



[11] Cabot, Chloé, Lina F. Soualmia, Badisse Dahamna, and Stéfan J. Darmoni. "SIBM at CLEF eHealth Evaluation Lab 2016: Extracting Concepts in French Medical Texts with ECMT and CIMIND." CLEF, 2016.

*CISMef Catalog and Index of French language Medical Sites. Available in: <http://www.chu-rouen.fr/cismef/>

** Pereira, Suzanne et al. (2008). Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. <http://ecmt.chu-rouen.fr/>



WHO-ART Ontology

WHO-ART : specialized ontology on Adverse Reaction Terminology

Example: searching **Fievre** (Fever) within WHO-ART with the annotator, in french or english

WHO-ART ontology (Main Class)

- ETAT GENERAL (Class) >> **DISOrder** (High level label)
- FIEVRE (Preferred Term), Term included (Synonyms):
 - FIEVRE D'ORIGINE MEDICAMENTEUSE
 - PYREXIE
 - REACTION FEBRILE

ECMT annotations of PRIMEGE notes

ECMT tool* extracted ontology concepts on PRIMEGE observations.csv

Basic annotation: [*Begin, End, Token, Label*, #Annotation, **Ontology**, Preferred Term]

Line 17: *d+ nuque et au dessus de l oreille , mieux avec advil , surveillance 3J . oreille nl -*

[[49,54,"advil","CHEM",1,"PHA","ADVIL"]]

Line 18: *se plaint de fluctuation d'anxiété -*

[[27,34,"anxiété","PHYS",1,"ART","ANXIETE"]]

Line 19: *rhinorrhée post, apyrexie pas de perte d'appétit - surveillance -*

[[0,9,"rhinorrhée","DISO",1,"ART","RHINORRHEE"],
[33,47,"perte d'appétit","DISO",2,"ICD","anorexie"]]

Term

>>

Preferred Term (synonym)

ECMT annotations of PRIMEGE notes

ECMT tool* extracted ontology concepts on PRIMEGE: 46422 notes with 1 413 030 tokens

		Ontology	Elements
• 3 Ontologies:	WHO- ART (Adverse Reaction Terminology)	ART	34334
	PHA (Racine des médicaments)	PHA	24225
	ICD-10 (International Statistical Classification of Diseases)	ICD	19451
		Overall	78010

- 7 categories (Label): ACTivities and Behaviors, **CHEMical** and drugs, **DISOrders**, PHENomena, PHYSiology, PROCedures, CONCepts and Ideas

Label	Annotations	Ontology
ACTI	272	Common
CHEM	24225	PHA
DISO	52403	Common
PHEN	295	Common
PHYS	796	Common
PROC	17	ICD
CONC	2	ICD
Overall	78010	5,5% of total tokens

CISMEF code: e.g. **ART**_IT_0171_IT5,
DISO label for term “sensation bizarre”



* Pereira, Suzanne et al. (2008). Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. <http://ecmt.chu-rouen.fr/>

3. Extracting ADR relation in Clinical Notes

Detection of relation between ADR and Medication in the sentence:

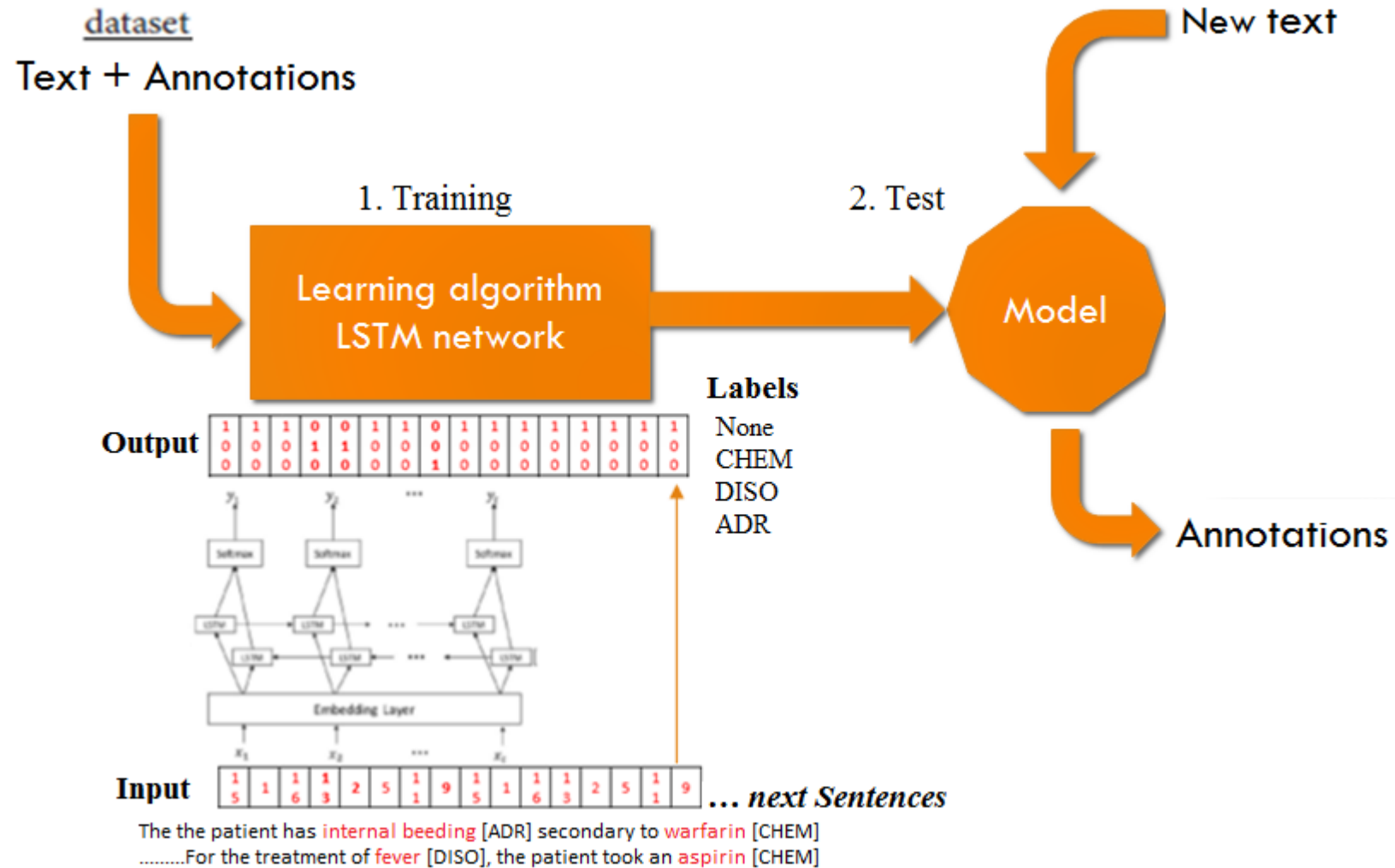
“the patient has internal bleeding (A) secondary to warfarin (B)”

[ADR]>>> dependency >>> [CHEM] Medication

Label for entity **A** is strongly related to the label prediction of **B**.

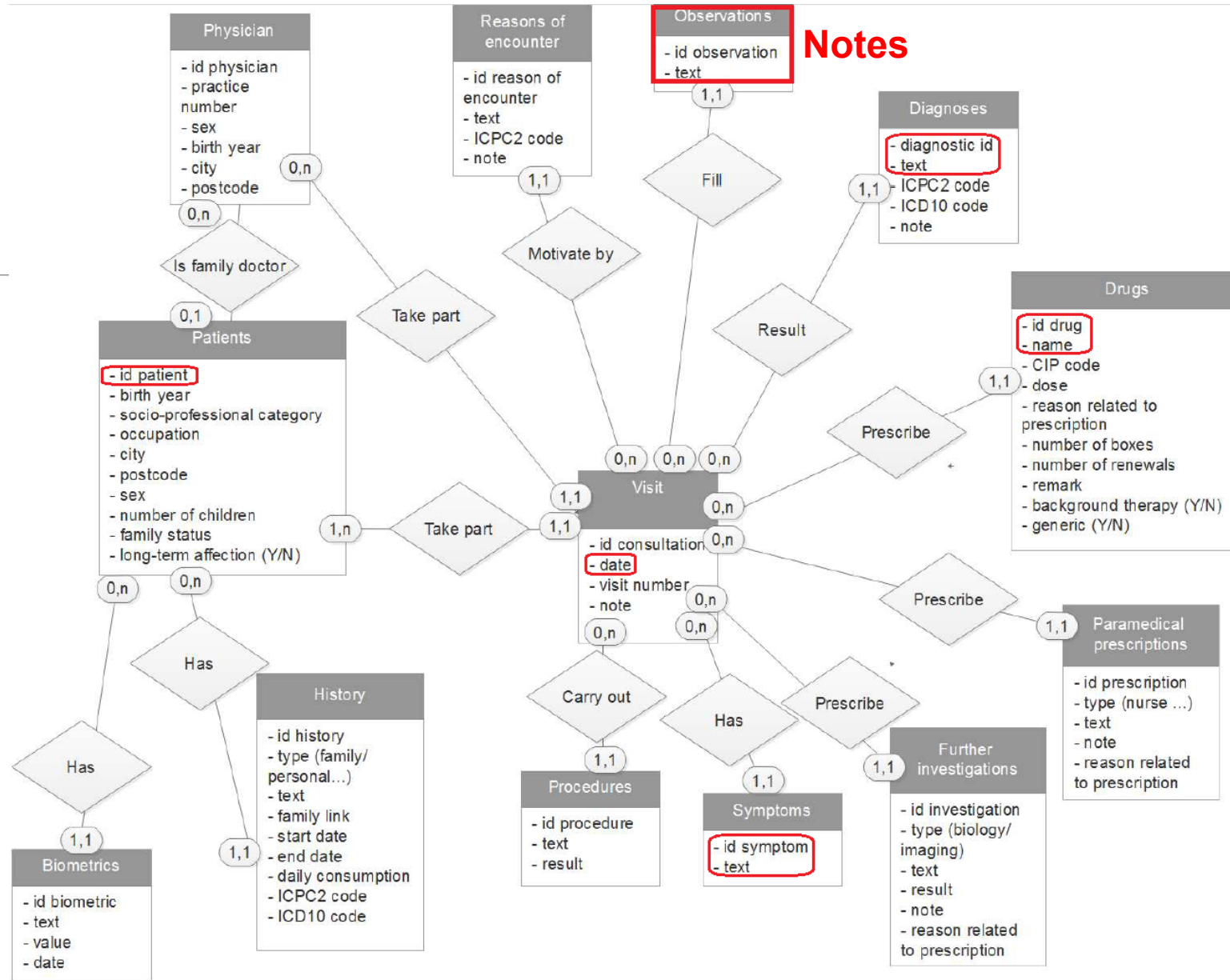
RNN (Recurrent Neural Network) models as **LSTM (Long Short Term Memory)** can model long term label dependencies [1].

3. Extracting ADR relation in Clinical Notes with LSTM



PRIMEGE database

Notes (id) are linked with the other elements of the database



Entity Association model centered on the visits*

*Lacroix-Hugues V., Darmon D., Pradier C. & Staccini P. "Creation of the first French database in primary care using the ICPC2: feasibility study." 16th World Congress on Medical and Health Informatics. 2017.

STATUS and FUTURE WORK

1. *First version of tool to extract clinical entities like disorders, medications, etc.*
2. *Obtain for each patient a time series with:*
 - *Notes*
 - *Medication*
 - *Diagnostic*
 - *Symptoms, etc*
3. *Build and train LSTM in order to detect ADR*



Named Entity Recognition approaches

Comparison of system for entity recognition when manually annotated data (from different sources) is used for evaluation [9]

Study	Reference	Method	Size	Recall	Precision	F-score
Leaman	[17]**	Lexicon-based (450 comments for system development)	3150	0.70	0.78	0.74
Nikfarjam	[8]**	Lexical pattern-matching (2400 comments). Association rule mining to identify patterns	1200	0.66	0.70	0.68
Nikfarjam	[3]	Supervised learning via Conditional Random Fields (CRFs)	1559	0.78	0.86	0.82
Jagannatha	[1]	Bi-LSTM-CRF (skip-crf-approx)	1154	0.83	0.81	0.82
Huynh	[5]*	CNNA (Convolutional Neural Network)	2972	0.84	0.82	0.83
Gurulingappa	[10]*	SVM (Support Vector Machines)	2972	0.86	0.89	0.87

*Systems using the same dataset or **subsets of the same dataset