

Traitement Automatique du Langage pour l'analyse des CR médicaux

04 juin 2025

Luca THIEBAUD & Adam REMAKI

adam.remaki@centralesupelec.fr

luca.thiebaud@centralesupelec.fr

■ Les CR médicaux à l'AP-HP

- Intégration « type » d'un CR médical
- Volumétrie et types de documents
- Enjeux du Traitement Automatique du Langage

■ EDS-NLP

- Présentation de la librairie
- Démo

■ Application 3 : Utiliser le TAL pour enrichir les données médicamenteuses



1.

Les Comptes-Rendus médicaux à l'AP-HP

■ Intégration « type » des CR médicaux

Unité de soins Intensifs de Cardiologie : 01 40 25 66 07
Unité de soins Intensifs de Cardiologie : Dect : 55033
Rendez-vous d'Echographie : 01 40 25 67 91
Rendez-vous de Consultation : 01 40 25 73 25
Rendez-vous d'Hospitalisation : 01 40 25 66 36

ASSISTANCE  HÔPITAUX
PUBLIQUE DE PARIS
BICHAT - CLAUDE BERNARD

Département de Cardiologie

N.I.P. : [REDACTED] Sexe : [REDACTED] Date de compte rendu : [REDACTED]
Patient : [REDACTED] Médecin : [REDACTED]
D.D.N. : [REDACTED] Interne : [REDACTED]
Hospitalisé du : [REDACTED] Ref : LR

COMPTE-RENDU D'HOSPITALISATION "GEME NORD"

Correspondant : [REDACTED] Rythmologie - BICHAT

Motif d'hospitalisation :
Implantation d'un Reveal.

Antécédents médicaux :
Syndrome du QT long congénital diagnostiqué en [REDACTED]
Epilepsie.
Dépression.
Rubéole.
Appendicectomie en [REDACTED]

Antécédents familiaux :
LQTS chez la mère, la grand-mère, 2 tantes/3 et une cousine.

Facteurs de risque cardiovasculaire :
Tabagisme sévère en [REDACTED]
Allergies : Amoxicilline, intolérance au Rivotril, au Dihydan, au Lamictal.

Traitement d'entrée :
Corgard 40mg/j.

Histoire de la maladie :
La patiente présente depuis l'âge de 10 ans environ, des syncopes à l'emporte-pièce liées à l'effort et aux émotions, initialement étiquetées d'épilepsie.
Vers l'âge de 20 ans apparaissent des crises un peu différentes des malaises précédents, avec confusion post critique et mouvements claudiques sans lien avec l'effort.
Par la suite, hospitalisation en rythmologie à Bichat avec diagnostic de syndrome de QT long congénital dans un contexte de malaise avec fibrillation atriale paroxystique. Le diagnostic est porté sur l'aspect électrocardiographique fortement évocateur ainsi qu'un test à l'isoprénaline. Les tests génétiques sont en attente.
La diagnostic d'épilepsie a donc été remis en cause et une décroissance du traitement anti épileptique a été instaurée entre [REDACTED], sans récédive de crise.

En ce qui concerne le LTQS, la patiente est traitée par Corgard 40mg/j depuis [REDACTED]

Récemment, elle a présenté 3 malaises mais sans perte de connaissance, avec sensation de palpitations et malaise général, durant 5mn, sans lien avec l'effort ou les émotions. Elle a consulté le Docteur DENJOY qui préconise l'implantation d'un Reveal.

Examen clinique à l'arrivée :
Pression artérielle 93/47 ; fréquence cardiaque 57/mn ; saturation 99% en air ambiant ; bruits du cœur réguliers sans souffle surajouté ; pas de signe d'insuffisance cardiaque gauche ni droite ; abdomen souple et indolore ; poumons clairs ; examen neurologique normal.

ECG :
Rythme sinusal régulier à 57/mn ; QRS fins ; axe normal ; QT corrigé mesuré à 470ms pour un RR 1080ms.

Biologie :

N° : 575803
Date du CR : [REDACTED]
N.I.P. : [REDACTED] Patient : FOS [REDACTED]
Page 1/2

Signé par : [REDACTED]

Sectionnement



Unité de soins Intensifs de Cardiologie : 01 40 25 66 07
Unité de soins Intensifs de Cardiologie : Dect : 55033
Rendez-vous d'Echographie : 01 40 25 67 91
Rendez-vous de Consultation : 01 40 25 73 25
Rendez-vous d'Hospitalisation : 01 40 25 66 36

ASSISTANCE  HÔPITAUX
PUBLIQUE DE PARIS
BICHAT - CLAUDE BERNARD

Département de Cardiologie

N.I.P. : [REDACTED] Sexe : [REDACTED] Date de compte rendu : [REDACTED]
Patient : [REDACTED] Médecin : [REDACTED]
D.D.N. : [REDACTED] Interne : [REDACTED]
Hospitalisé du : [REDACTED] Ref : LR

COMPTE-RENDU D'HOSPITALISATION "GEME NORD"

Correspondant : [REDACTED] Rythmologie - BICHAT

Motif d'hospitalisation :
Implantation d'un Reveal.

Antécédents médicaux :
Syndrome du QT long congénital diagnostiqué en [REDACTED]
Epilepsie.
Dépression.
Rubéole.
Appendicectomie en [REDACTED]

Antécédents familiaux :
LQTS chez la mère, la grand-mère, 2 tantes/3 et une cousine.

Facteurs de risque cardiovasculaire :
Tabagisme sévère en [REDACTED]
Allergies : Amoxicilline, intolérance au Rivotril, au Dihydan, au Lamictal.

Traitement d'entrée :
Corgard 40mg/j.

Histoire de la maladie :
La patiente présente depuis l'âge de 10 ans environ, des syncopes à l'emporte-pièce liées à l'effort et aux émotions, initialement étiquetées d'épilepsie.
Vers l'âge de 20 ans apparaissent des crises un peu différentes des malaises précédents, avec confusion post critique et mouvements claudiques sans lien avec l'effort.
Par la suite, hospitalisation en rythmologie à Bichat avec diagnostic de syndrome de QT long congénital dans un contexte de malaise avec fibrillation atriale paroxystique. Le diagnostic est porté sur l'aspect électrocardiographique fortement évocateur ainsi qu'un test à l'isoprénaline. Les tests génétiques sont en attente.
La diagnostic d'épilepsie a donc été remis en cause et une décroissance du traitement anti épileptique a été instaurée entre [REDACTED], sans récédive de crise.

En ce qui concerne le LTQS, la patiente est traitée par Corgard 40mg/j depuis [REDACTED]

Récemment, elle a présenté 3 malaises mais sans perte de connaissance, avec sensation de palpitations et malaise général, durant 5mn, sans lien avec l'effort ou les émotions. Elle a consulté le Docteur DENJOY qui préconise l'implantation d'un Reveal.

Examen clinique à l'arrivée :
Pression artérielle 93/47 ; fréquence cardiaque 57/mn ; saturation 99% en air ambiant ; bruits du cœur réguliers sans souffle surajouté ; pas de signe d'insuffisance cardiaque gauche ni droite ; abdomen souple et indolore ; poumons clairs ; examen neurologique normal.

ECG :
Rythme sinusal régulier à 57/mn ; QRS fins ; axe normal ; QT corrigé mesuré à 470ms pour un RR 1080ms.

Biologie :

N° : 575803
Date du CR : [REDACTED]
N.I.P. : [REDACTED] Patient : FOS [REDACTED]
Page 1/2

Signé par : [REDACTED]

■ Intégration « type » des CR médicaux

Pseudonymisation et
extraction du texte



Correspondant : [REDACTED] - Rythmologie - BICHAT

Motif d'hospitalisation :
Implantation d'un Reveal.

Antécédents médicaux :
Syndrome du QT long congénital diagnostiqué en [REDACTED]
Epilepsie.
Dépression.
Rubéole.
Appendicectomie en [REDACTED]

Antécédents familiaux :
LQTS chez la mère, la grand-mère, 2 tantes/3 et une cousine.

Facteurs de risque cardiovasculaire :
Tabagisme sevré en [REDACTED]
Allergies : Amoxicilline, intolérance au Rivotril, au Dihydan, au Lamictal.

Traitement d'entrée :
Corgard 40mg/j.

Histoire de la maladie :
La patiente présente depuis l'âge de 10 ans environ, des syncopes à l'emporte-pièce liées à l'effort et aux émotions, initialement étiquetées d'épilepsie.
Vers l'âge de 20 ans apparaissent des crises un peu différentes des malaises précédents, avec confusion post critique et mouvements claudiques sans lien avec l'effort.
Par la suite, hospitalisation en rythmologie à Bichat avec diagnostic de syndrome de QT long congénital dans un contexte de malaise avec fibrillation atriale paroxystique. Le diagnostic est porté sur l'aspect électrocardiographique fortement évocateur ainsi qu'un test à l'isoprénaline. Les tests génétiques sont en attente.
La diagnostic d'épilepsie a donc été remis en cause et une décroissance du traitement anti épileptique a été instaurée entre [REDACTED], sans récurrence de crise.

En ce qui concerne le LQTS, la patiente est traitée par Corgard 40mg/j depuis [REDACTED]

Récemment, elle a présenté 3 malaises mais sans perte de connaissance, avec sensation de palpitations et malaise général, durant 5mn, sans lien avec l'effort ou les émotions. Elle a consulté le Docteur DENJOY qui préconise l'implantation d'un Reveal.

Examen clinique à l'arrivée :
Pression artérielle 93/47 ; fréquence cardiaque 57/mn ; saturation 99% en air ambiant ; bruits du cœur réguliers sans souffle surajouté ; pas de signe d'insuffisance cardiaque gauche ni droite ; abdomen souple et indolore ; poumons clairs ; examen neurologique normal.

ECG :
Rythme sinusal régulier à 57/mn ; QRS fins ; axe normal ; QT corrigé mesuré à 470ms pour un RR 1080ms.

Biologie :
N° : [REDACTED] Signé par : [REDACTED]
Date du CR : [REDACTED]
N.L.P. : [REDACTED] Patient : FOS # [REDACTED]
Page 1/2

Motif d'hospitalisation:

Implantation d'un Reveal.

Antécédents médicaux:

Syndrome du QT long congénital diagnostiqué en DD/MM/YYYY

Epilepsie

Dépression

Rubéole

Appendicectomie en DD/MM/YYYY

Antécédents familiaux:

LQTS chez la mère, la grand-mère, 2 tantes/3 et une cousine

Facteurs de risque cardiovasculaire:

Tabagisme sevré en DD/MM/YYYY

Allergies: Amoxicilline, intolérance au Rivotril, au Dihydans, au Lamictal

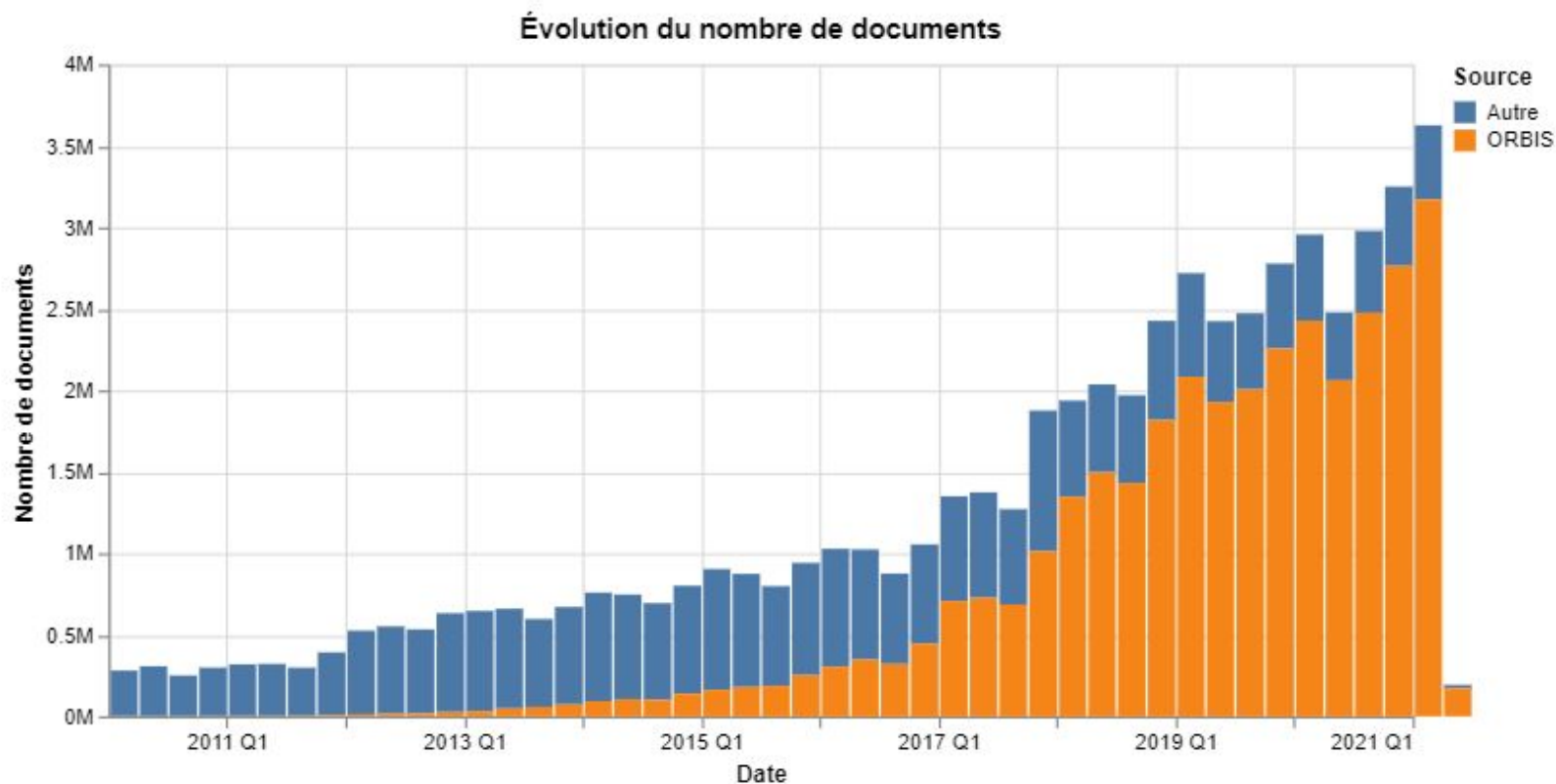
Traitement d'entrée:

Cograd 40mg/j

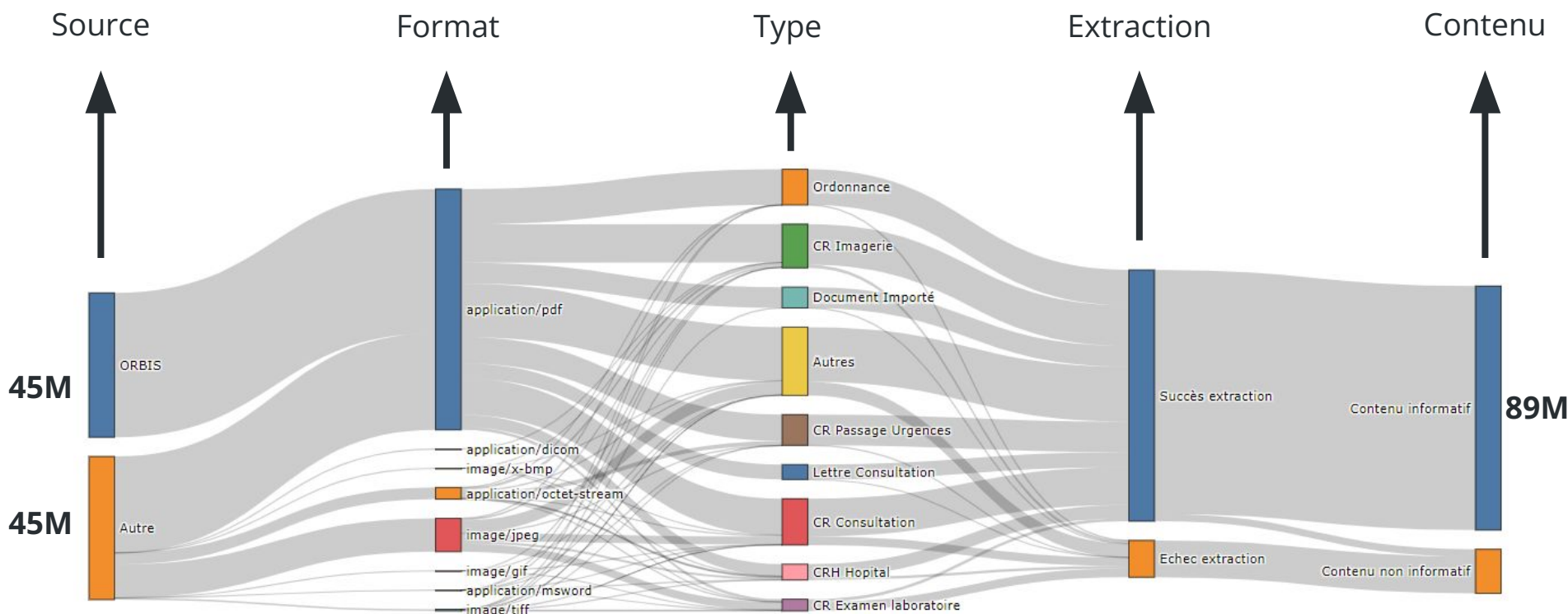
Histoire de la maladie:

...

■ Volumétrie et types de documents



■ Volumétrie et types de documents



Intégration des documents textuels dans l'EDS

■ Enjeux du Traitement Automatique du Langage (= TAL = NLP)

□ CR médicaux extrêmement riches et informatifs

- *Antécédents*
- *Comorbidités*
- *Diagnostics*
- *Traitements proposés et administrés*

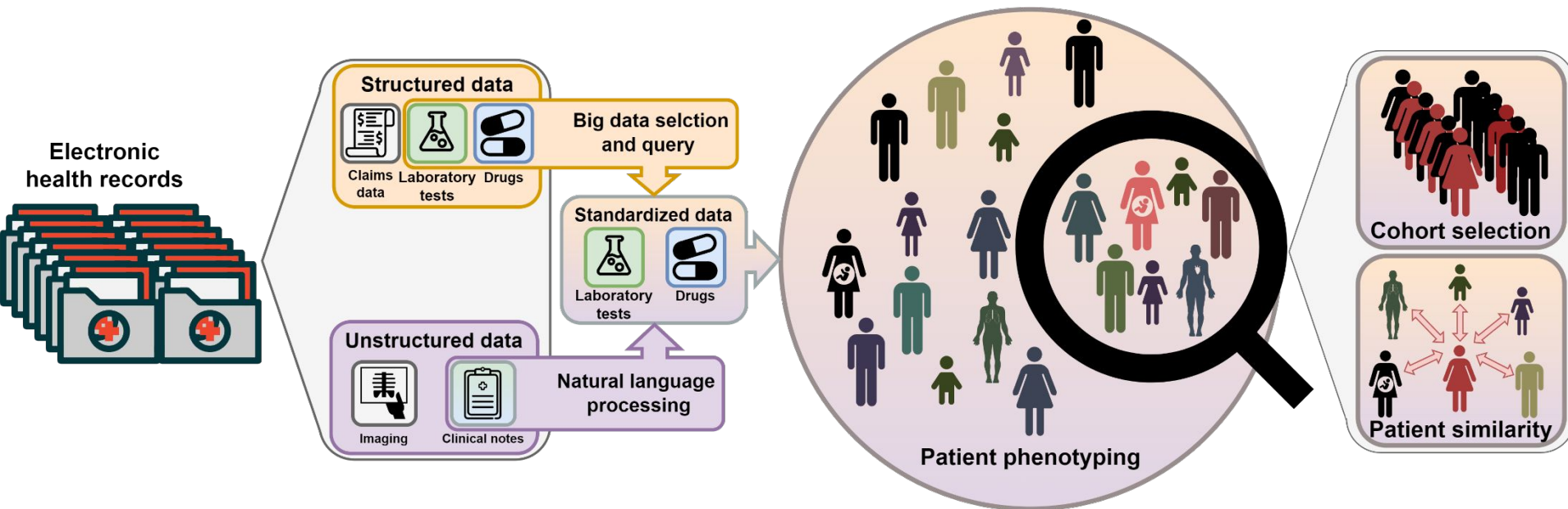
□ Mais assez difficiles à exploiter

- *Beaucoup d'acronymes*
- *Phrases assez pauvres d'un point de vue syntaxique*
- *Vocabulaire très diversifié*

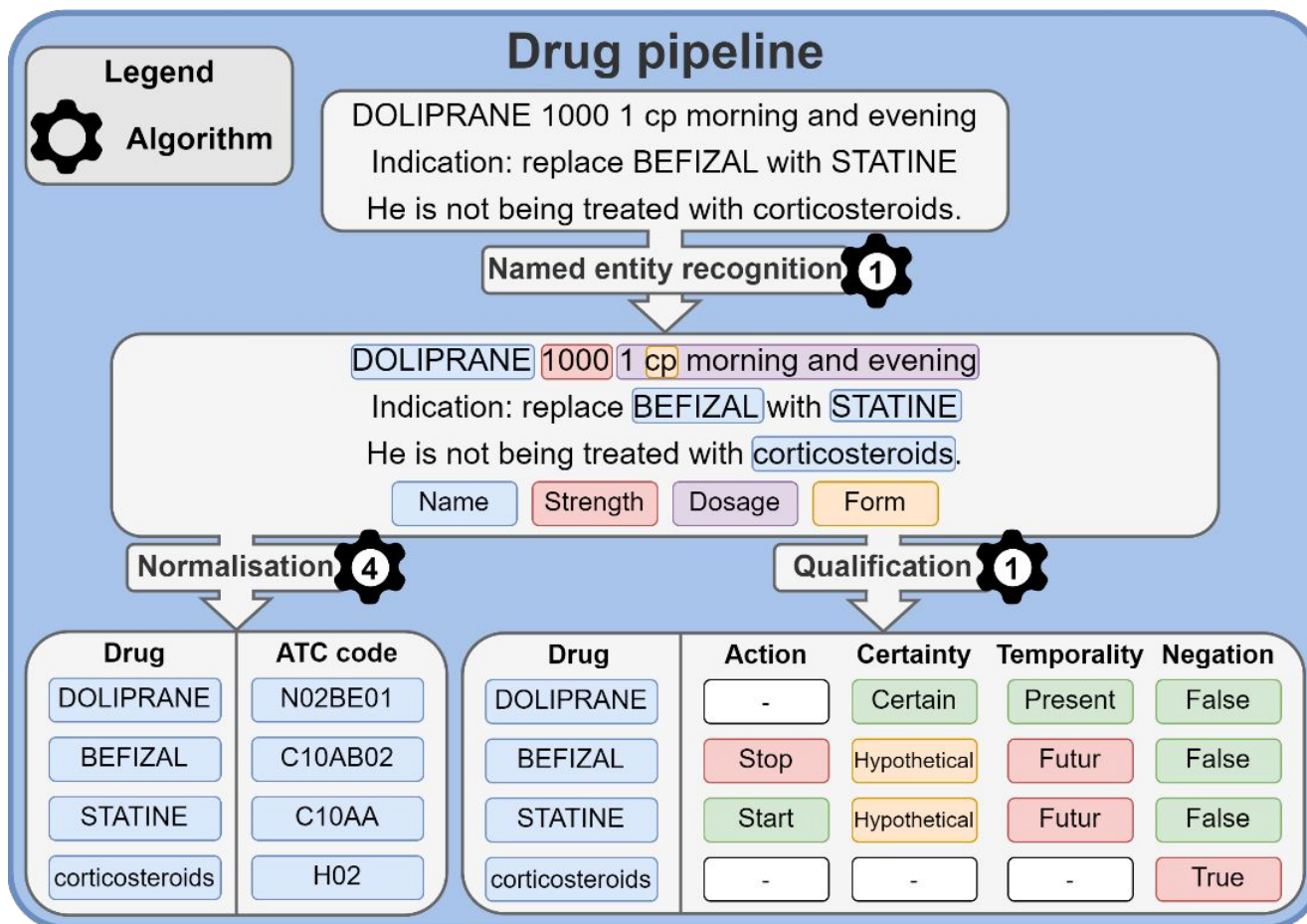
□ Avec un besoin de qualifier précisément les entités

- « Le patient est/n'est pas/semble être diabétique »
- « Le père du patient est diabétique »
- « Le patient dit avoir fait une TS il y a un mois »

Objectif



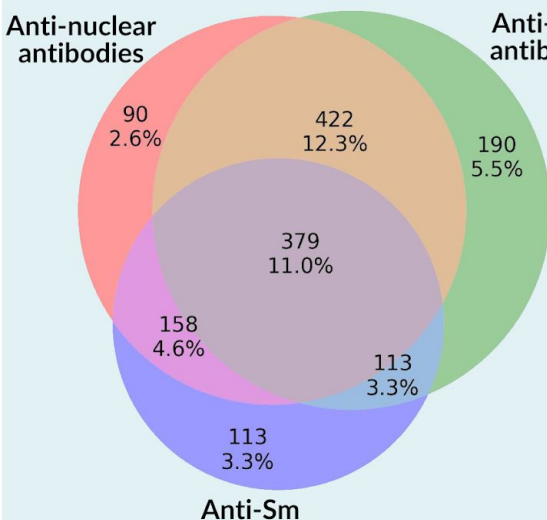
Méthode



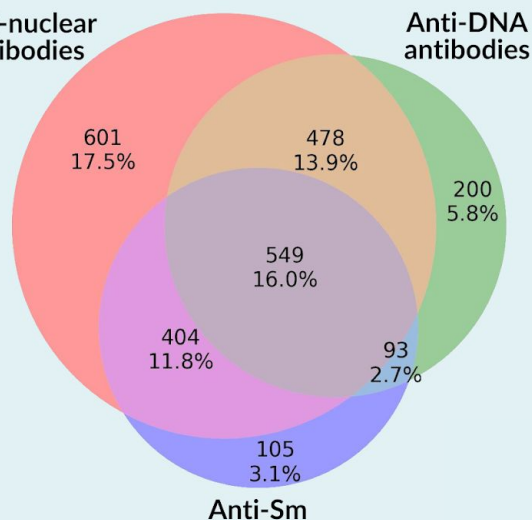
Résultats

3437 patients with lupus

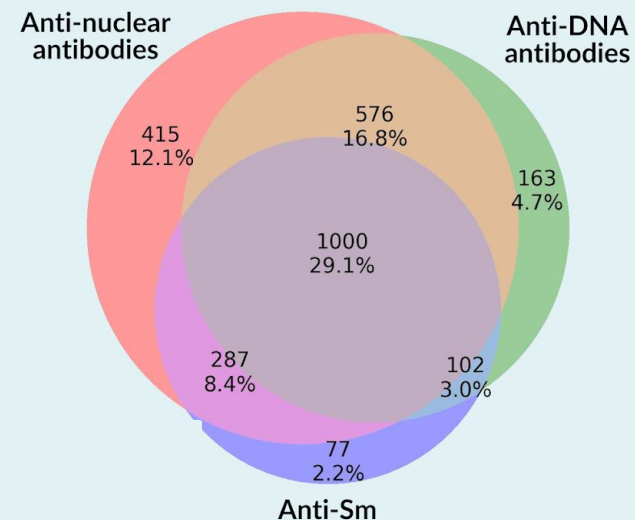
Structured data
1465 patients (42.6 %)



Unstructured data
2430 patients (70.7 %)



Structured + unstructured data
2620 patients (76.2 %)





2.

EDS-NLP : Librairie de TAL
sur le texte clinique

■ Objectif : faciliter l'exploitation des documents médicaux

- Proposer un cadre simple de traitement des données textuelles
- Mettre à disposition des *baselines* **performantes** et **validées**
- Simplifier le partage des algorithmes de TAL (pour les autres projets & relecteurs)
- Accent sur la reproductibilité

■ En pratique

- Librairie Python
- Basée sur la librairie spaCy (spécialisée en TAL sur de larges volumes de textes)
- Sous forme de *pipelines* juxtaposables:
 - *Sentences*
 - *Normalizer*
 - *Matcher (Regex vs PhraseMatcher)*
 - *Négation/Parenté/Hypothèse*

■ Démo !



3.

Application 3

■ Pull le repo GitHub pour mettre à jour le dossier

- Lien URL : <https://github.com/Aremaki/edstuto>
- Si edsnlp ne s'installe pas, aller sur Google Collab:
 - *Importer le notebook*
 - *Importer le dossier data*
 - *Importer le fichier viz*

■ Ouvrir le notebook **exercises/exercise-3**

- Contexte : Utiliser les méthodes de TAL pour enrichir les données structurées sur les médicaments
- Objectif : comparaison des courbes de survie avec et sans NLP pour extraire le médicament administré (drugA/drugB/control)
- Plan d'action :
 - *Prise en main de la table df_note*
 - *Construction d'un pipeline d'extraction d'entité médicamenteuse*
 - *Tracé des courbes de survie avec et sans enrichissement des données médicamenteuses par NLP*



4.

Projet

■ Objectifs du jour

- Prise en main de la table *df_note*
- Lecture de quelques CR pour identifier les enrichissements potentiels
- Construction d'un pipeline NLP ad-hoc
- Conversion des résultats en données tabulaires
- Validation des performances de l'algorithme avec de l'annotation

■ Données

- Base de données disponible dans le dossier *final_project/data*

■ Reproductibilité : vos notebooks seront lancés sur un set de données de test non fourni et doivent fonctionner tels quels.

■ Compétences évaluées :

- **Compétence C4** : Avoir le sens de la création de valeur
- **Compétence C6** : Être opérationnel, responsable et innovant dans le monde numérique
- **Compétence C7** : Savoir convaincre
- **Compétence C8** : Mener un projet en équipe