Semantic Clinical Data Warehouse in Rouen University Hospital, Normandy, France

Prof. Stéfan Darmoni, MD, PhD

Badisse Dahamna, Romain Lelong, Ivan Kergourlay, Kevin Billey, Julien Grosjean

Department of Biomedical Informatics (DBI), Rouen University Hospital, Normandy, France
&
LIMICS INSERM U1142, Sorbonne University, France

D2IM
Département d'Informatique
et d'Information Médicales

RUH DBI

- Rouen University Hospital, Normandy
- N =15 (vs. 100 in Harvard Medical School)
  - Heterogeneity of competence ; double competence
  - Two medical informaticians
  - Five engineers
  - Three medical librarians (CISMeF, LiSSa)
  - Two/three medical residents in public health + one GP resident
  - Four PhD applicants

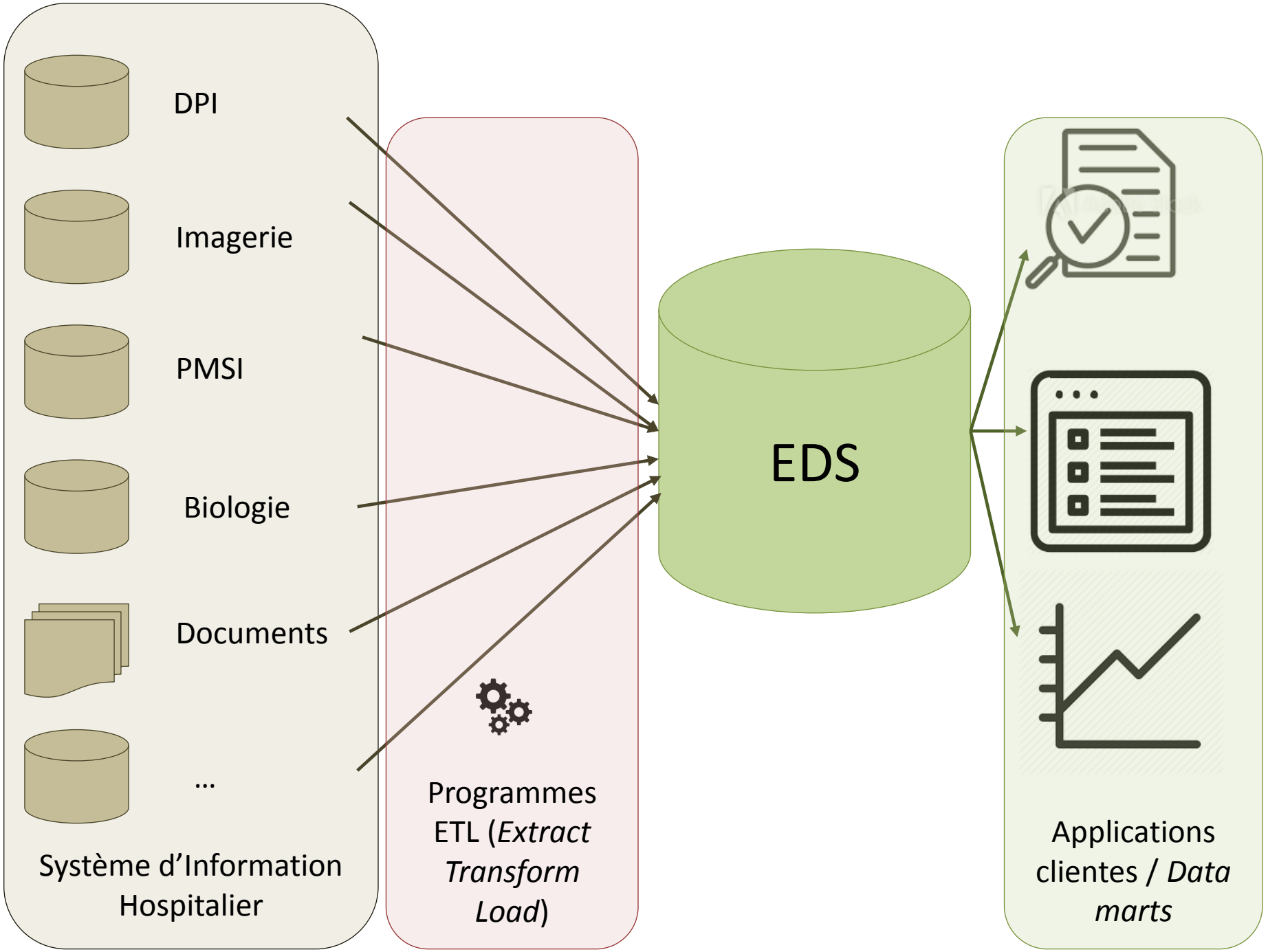
In 2019, RUH DBI became part of the LIMICS, main French lab in biomedical informatics

LIMICS

Definition

- Health Data Warehouse (HDW) are computer tools allowing collection, integration, then data mining. These data are extracting from various clinical information systems: EHR, LIS, RIS/PACS, CPOE, etc…
- Agregation of a maximum of data available from patients
  - First step, HDW from DRG data
- No matter what is the Clinical Information System (CIS)
  - As far as possible, retrieving all the data from CIS, no matter old they are; going back to 2000 ≠ Paris Hospitals (APHP)

➔ Allowing to link information and to select as precisely as possible the right patients


DPI
Imagerie
PMSI
Biologie
Documents
…
Système d'Information Hospitalier
Programmes ETL (*Extract Transform Load*)
EDS
Applications clientes / *Data marts*

HDW Objectives

- To optimize DRGs by semiautomatic detection of atypical profiles between coding and HDW data (DBI)
- To Improve clinical research thanks to feasibility studies prior to clinical trials & optimization of inclusions
- Detection specific patients profiles
  - e.g. patients frequently admitted to the emergency department
- To create and maintain epidemiological cohorts and registries
- To detect specific adverse events
  - Vigilances, iatrogeny, cross infections
- To create and follow-up quality indicators
- To Develop and assess computer aided decision support systems (CDASS)

État des lieux des EDS

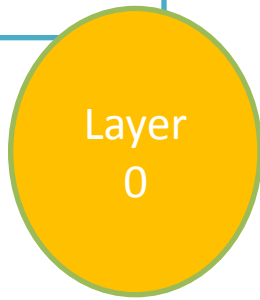
Dans le monde :

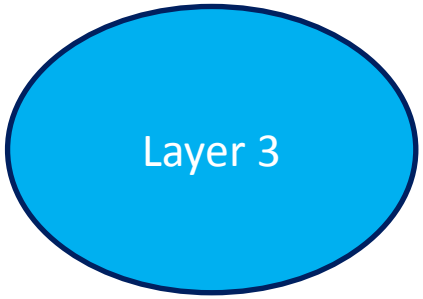
- i2b2 (Harvard MS)
- …

En France :

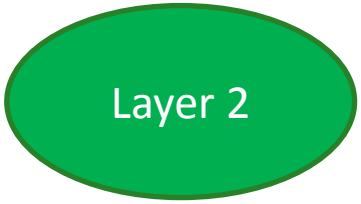
- ehop (CHU de Rennes)
- Dr Warehouse (Necker, Foch – Paris)
- ConSoRe
  Continuum Soins Recherche
- …

Semantic HDW based on three independent layers

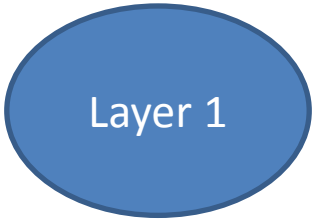




Multiterminology Multilingual Semantic search engine



ECMT Semantic Annotator (NLP & Deep Learning)



HeTOP Terminology server

NoSQL Architecture

HeTOP Terminology server

Layer 1

- Multi Terminology & cross lingual → matrix navigation (among languages & among terminologies
- 86 termino-ontologies included in 32 languages
- 2 M concepts in English; 1.2 M in French
  - ≈ 145 K concepts in French in UMLS (2018AB) vs. ≈ 445 K in HeTOP (x2.98)
- Over 100 million RDF triplets (2014) → big data +++

ECMT Semantic Annotator (NLP & Deep Learning)

Layer 2

- Based on HeTOP; 50 chosen KOS out of 75 (no interface terminologies)
- 11.8 M health documents
- Processing time: 22-24 hours (two servers 1 To; one with 196 cores and the second with 144 cores
- 5.2 G medical concepts extracted; 2.6 G after filtering

Siefridt C, Grosjean J, Lefebvre T, Rollin L, Darmoni S, Schuers M. Evaluation of automatic annotation by a multi-terminological concepts extractor within a corpus of data from family medicine consultations. Int J Med Inform. 2020 Jan;133:104009. doi: 10.1016/j.ijmedinf.2019.104009. Epub 2019 Nov 1.

Neveol & coll. Clinical Natural Language Processing in languages other than English: opportunities and challenges. Journal of Biomedical Semantics 2018 9:12

Overall ECMT process

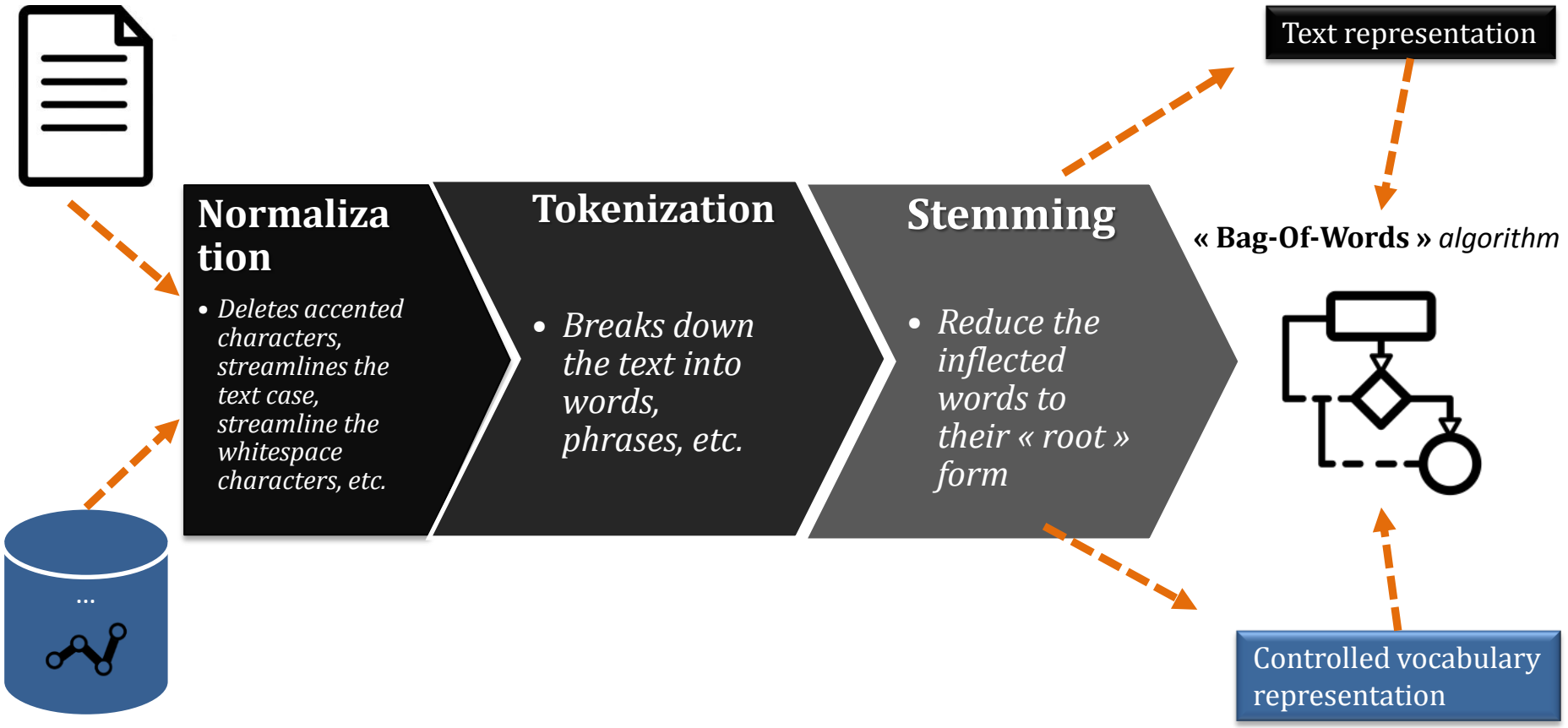


Table 7. System performance for ICD10 coding on the **French raw** test corpus in terms of Precision (P), recall (R) and F-measure (F). A horizontal dash line places the frequency baseline performance. The top part of the table displays official runs, while the bottom part displays non-official and baseline runs.

	ALL				EXTERNAL			
	Team	P	R	F	**Team**	P	R	F
Official runs	SIBM-run1	**.857**	**.689**	**.764**	SIBM-run1	**.567**	**.431**	**.490**
	LITL-run2	.666	.414	.510	LIRMM-run1	.443	.367	.401
	LIRMM-run1	.541	.480	.509	LIRMM-run2	.443	.367	.401
	LIRMM-run2	.540	.480	.508	LITL-run2	.560	.283	.376
	LITL-run1	.651	.404	.499	LITL-run1	.538	.277	.365
	TUC-MI-run2	.044	.026	.033	TUC-MI-run2	.010	.004	.005
	TUC-MI-run1	.025	.015	.019	TUC-MI-run1	.006	.005	.005
	average	.475	.358	.406	**average**	.367	.247	.292
	median	.541	.414	.508	**median**	.443	.283	.376
Non-official	LIMSI-run2	.872	.784	.825	LIMSI-run2	.700	.594	.643
	LIMSI-run1	.883	.760	.817	LIMSI-run1	.709	.559	.625
	TUC-MI-run1-corrected	.883	.539	.669	TUC-MI-run1-corrected	.780	.290	.423
	TUC-MI-run2-corrected	.882	.536	.667	TUC-MI-run2-corrected	.767	.283	.414
	UNIPD-run1	.629	.468	.537	UNIPD-run2	.350	.381	.365
	UNIPD-run2	.518	.384	.441	UNIPD-run1	.362	.251	.296
	Mondeca-run1	.375	.131	.194	Mondeca-run1	.335	.228	.271
	Frequency baseline	.339	.237	.279	Frequency baseline	.381	.110	.170

Multiterminology Multilingual Semantic search engine

Layer 3

- First step, definition of a model as light and as compact as possible
- Search engine based on Semantic Web
  - Explosion (subsumption) based on hierarchy at a multiterminology level
  - Other relations may be used: semantic expansions based on inter-terminology semantic mappings
  - Multilingual +++

Lelong, Romain; Soualmia, Lina F; Grosjean, Julien; Taalba, Mehdi & Darmoni, SJ. Building a Semantic Health Data Warehouse: Evaluation of a search tool in Clinical trials. JMIR Medical Informatics , 2019 (in press).

Multiterminology Multilingual Semantic search engine

Layer 3

- Very generic tool
- Designed to be used for N patients (regular use of HDW) or for one patient (not politically correct)

→ use of HDW to consult patient data for CARE +++

Two or three generations of software (LIS, RIS, EHR) compiled in one HDW

Show me all the electroretinography reports for this patient

→ one second vs. at least one minute

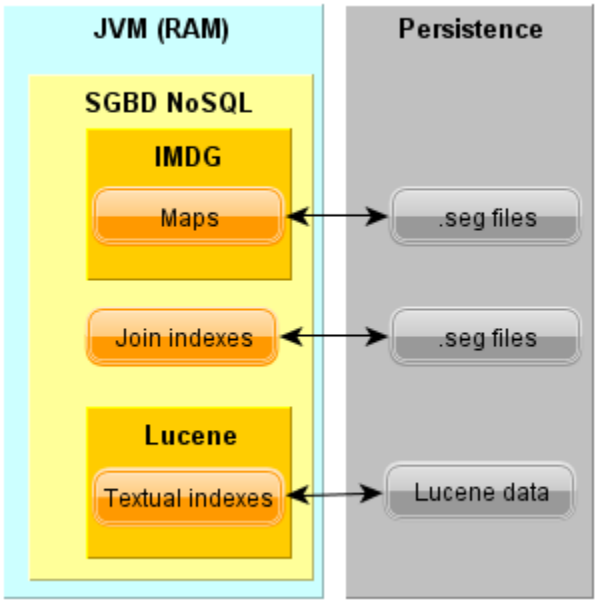
TECHNICAL ASPECTS

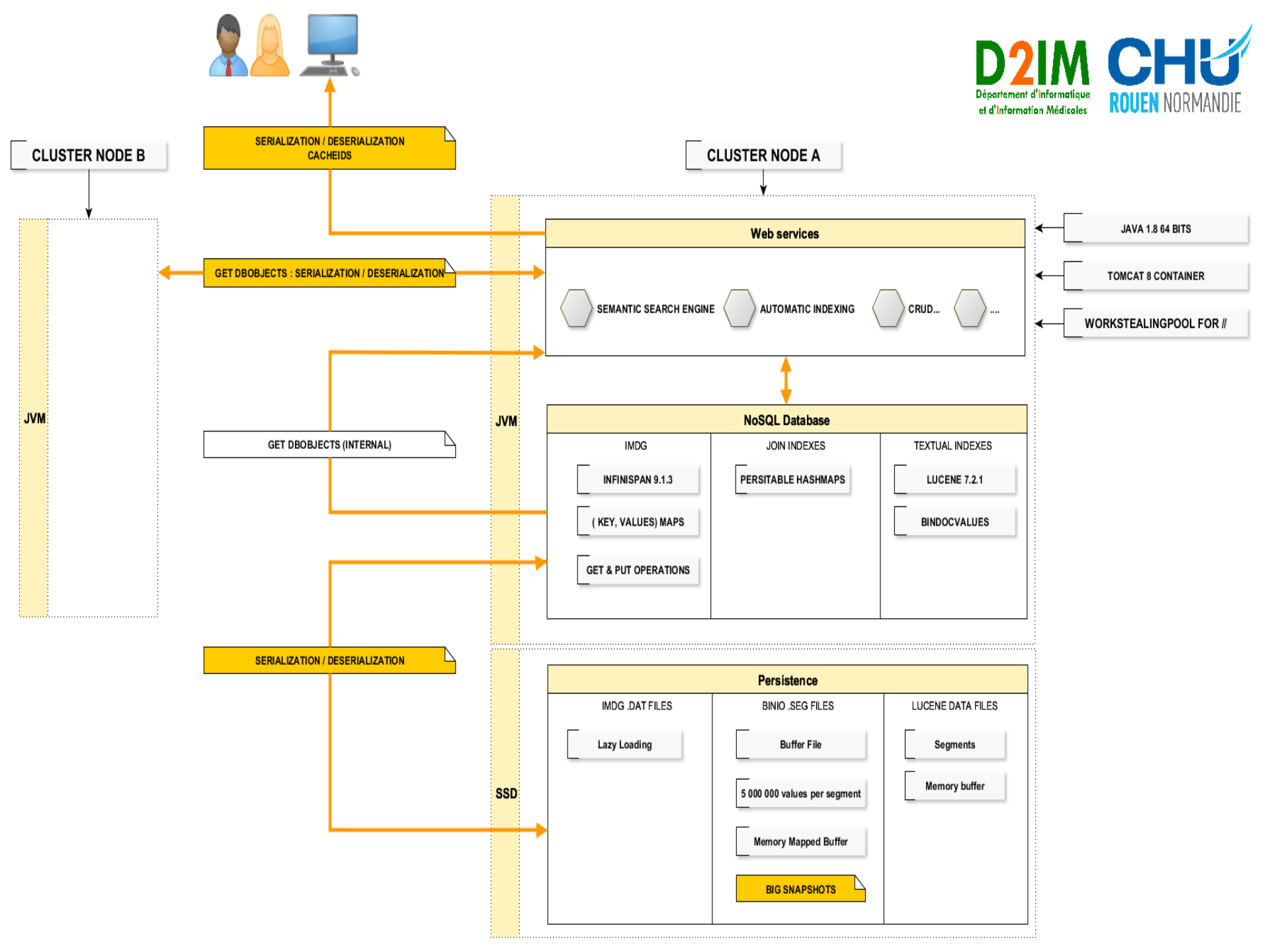
NoSQL Architecture

Layer 0

- Java
- Custom NoSQL (In Memory Data Grid – IMDG-): Key-Value store, join indexes (replacing SQL joins) & Lucene (NLP)

- Three powerful servers:
  - 1 To RAM & 196 cores (prod)
  - 1 To RAM & 144 cores (preprod) * 2
  - Semantic Annotator + deep learning




D2IM
Département d'Informatique et d'Information Médicales
CHU ROUEN NORMANDIE
CLUSTER NODE B
SERIALIZATION / DESERIALIZATION CACHEIDS
CLUSTER NODE A
JVM
GET DBOBJECTS : SERIALIZATION / DESERIALIZATION
Web services
JAVA 1.8 64 BITS
TOMCAT 8 CONTAINER
WORKSTEALINGPOOL FOR //
SEMANTIC SEARCH ENGINE
AUTOMATIC INDEXING
CRUD...
....
JVM
GET DBOBJECTS (INTERNAL)
NoSQL Database
IMDG
INFINISPAN 9.1.3
(KEY, VALUES) MAPS
GET & PUT OPERATIONS
JOIN INDEXES
PERSITABLE HASHMAPS
TEXTUAL INDEXES
LUCENE 7.2.1
BINDOCVALUES
SERIALIZATION / DESERIALIZATION
SSD
Persistence
IMDG .DAT FILES
Lazy Loading
BINIO .SEG FILES
Buffer File
5 000 000 values per segment
Memory Mapped Buffer
BIG SNAPSHOTS
LUCENE DATA FILES
Segments
Memory buffer

HEALTH DATA WAREHOUSE - SECURE ARCHITECTURE PROPOSAL

HDW PRODUCTION ENVIRONMENT

SECURE HOSTING

HDW APPLICATION SERVER (HTTPS)

- NO NOMINATIVE DATA
- HDW INTERNAL PATIENTS IDS
- ENCRYPTED PARTITION
- ANONYMIZED RECORDS

IMDG

MEDICAL DATA PERSISTENCE SERVER

- NO NOMINATIVE DATA
- HDW INTERNAL PATIENTS IDS
- ENCRYPTED PARTITION
- ANONYMIZED RECORDS

POSTGRESQL 10

NOMINATIVE DATA PERSISTENCE SERVER

- ISOLATED NOMINATIVE INFORMATIONS
- NO MEDICAL DATA
- ENCRYTED JOIN (SALT)
  HDW PATIENTS IDS - HOSPITAL IDS
- ENCRYPTED PARTITION
- ENCRYPTION OF SENSITIVE DATA (NAMES..)
  IN DATABASE

POSTGRESQL 10

RESTRICTION TO AUTHORIZED IP

USERS

REFERENT USER

AUTHENTIFIED SESSION (BROWSER)

D2IM (DIM+SIBM)

SECURE DELIVERY (TEMPORARY)
ENRCYPTED ARCHIVE

HDW INTERNAL NEEDS (APPROVED DEMAND)

SSH

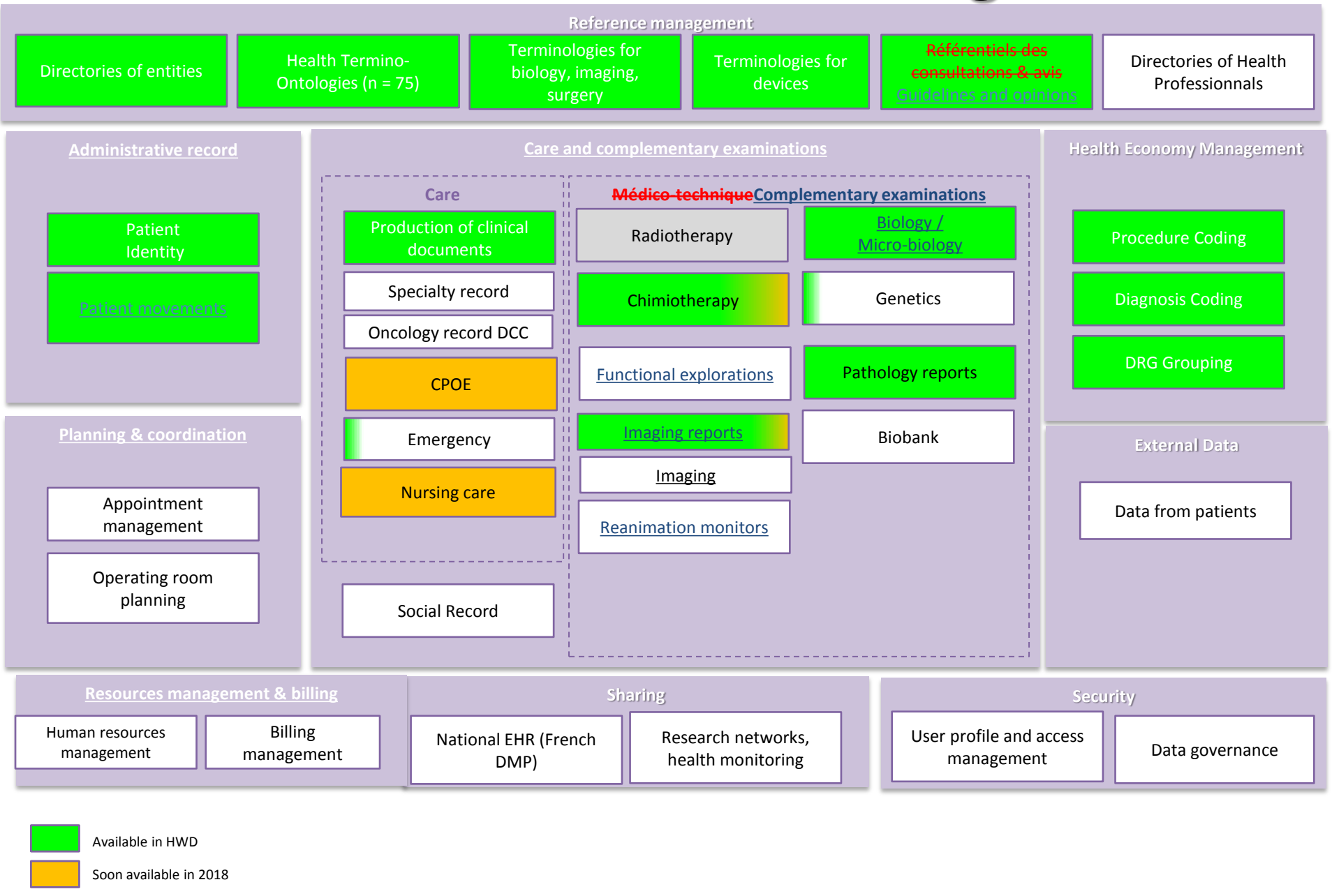
HDW ADMINISTRATORS

Results

HDW Rouen - volumetry

- 1.86 M patients
- 13.28 M stays (H – C)
- 13.4 M documents (since 2000)
  - 5 G extracted medical concepts (ECMT)
- 121.1 M unitary biology tests (Na, K) (since 2004)
- DRG: 9.4 M diagnoses; 8.6 M procedures

HDW Rouen – functional coverage



Main steps are already performed…

- A team of FIVE engineers are devoted to this project from DBMI + physicians from the DRG
- Scalability of the French semantic annotator
  - 11.8 M reports extracted in less then 24h; possible to rerun again every week if necessary (6ms for one report)
- Scalability of the semantic search engine
  - Using a NoSQL architecture and two powerful servers, response time ≈ 2 s
  - Might be used in care context
    - Display the last US of this patient ASAP

Table 1 - List of different terminologies. with the number of annotations (before & after filtering)

Terminologies (SOC)	Number of annotations (after filtering)	Number of annotations (before filtering)	Filtering ratio
SNOMED CT	394 133 994	881 884 314	2.2
NCIt	319 853 952	843 195 067	2.6
MeSH	295 537 298	1 024 585 229	3.5
SNOMED 3.5	219 706 745	440 228 408	2.0
French Public Health Thesaurus	179 747 539	454 354 922	2.5
MedDRA	137 653 806	225 100 880	1.6
TUV	106 616 463	171 231 559	1.6
RADLEX	80 197 479	150 406 338	1.9
FMA	55 350 010	123 777 281	2.2
CISMeF	51 051 547	138 204 239	2.7

Where ICD10 & CCA –French ICHI-?

Top 10 of Terminologies using unique annotated concepts (after filtering)

Name of KOS	Total number of concepts	Number of translated concepts (%)	Number of unique identified concepts	Terminology coverage (%)
SNOMED CT Concept	326946	194611 (59.5)	**59330**	30.5
Notion SNOMED	100908	100908 (100)	36229	35.9
NCIt Concept	93925	68776 (73.2)	25315	36.8
MedDRA	44226	44226 (100)	22711	51.4
MeSH Descriptor	28329	28329 (100)	18288	**64.6**
MedDRA PT	21612	21612 (100)	13580	**62.8**
MeSH Concept	365731	102116 (27.9)	12625	12.4
FMA	81041	16629 (20.5)	8084	48.6
Radlex	42313	10259 (24.2)	6114	59.6
French Public Health Thesaurus	7087	7087 (100)	6089	**85.9**

HeTOP | HeTOP | ((("Partial color blindnes | Abderhalden-Fanconi - F | grouping - Traduction fr | DRG grouping - Recherc | deuteranopie - Re

Non sécurisé | https://cisu4533.adm.chu-rouen.lan:8091/asis/

ASIS

Accès Sémantique à l'Information de Santé

Type d'entité recherchée :

Patient(s) | Hospitalisation(s) | Sejour(s) | Prise(s) en charge en UF | Analyse(s) biologique(s) | Diagnostic(s) | Acte(s) | Compte(s)-rendu(s)

	Patient(s)	Sexe	Homme / Femme / Autre	
ET	Diagnostic(s)	Terminologie(s)	...	Commencer à saisir
ET	Analyse(s) biologique(s)	Type de l'analyse	...	Commencer à saisir
ET	Sejour(s)	Date d'entrée dans l'épisc	=	2018-04-30

Ajouter une ligne

Construction de la requete :

Exemple de Requetes - Requête en language moteur

ASIS

Accès Sémantique à l'Information de Santé

Searched entity type :

Patient | Hospitalization | Stay | Patient managment | Biological test | Diagnosis | Procedure | Record

	Patient	Gender	Male / Female / Other	
ET	Diagnosis	Terminology(ies)	...	Enter terms
ET	Biological test	Type of biological test	...	Enter terms
ET	Stay	undefined	=	2018-04-30

Ajouter une ligne

Construction de la requete :

(@1 PATIENTS Male 105696)

Exemple de Requetes - Requête en language moteur

of concept provides Information Retrieval capabilities among a data subset of the health data warehouse of the university hospital of Rouen. Access to these data is regulated and an official request must be made at the

Health Data warehouse devoted to University Hospital of Rouen.

ASIS

Accès Sémantique à l'Information de Santé

Type d'entité recherchée :

Patient(s) | Hospitalisation(s) | Sejour(s) | Prise(s) en charge en UF | Analyse(s) biologique(s) | Diagnostic(s) | Acte(s) | Compte(s)-rendu(s)

	Patient(s)	Sexe	Homme / Femme / Autre				
ET	Diagnostic(s)	Terminologie(s)	...	J45 asthme	Expl	Source de données	Type
ET	Analyse(s) biologique(s)	Type de l'analyse	...	Commencer à saisir			
ET	Sejour(s)	Date d'entrée dans l'épisc	=	2018-04-30			

Ajouter une ligne

Construction de la requete :

(@1 PATIENTS Homme 105696) ET (@2 DIAGNOSTICS J45 asthme... 56)

Exemple de Requetes - Requête en language moteur

ASIS

Accès Sémantique à l'Information de Santé

Type d'entité recherchée :

Patient(s) | Hospitalisation(s) | Sejour(s) | Prise(s) en charge en UF | Analyse(s) biologique(s) | Diagnostic(s) | Acte(s) | Compte(s)-rendu(s)

	Patient(s)	Sexe	Homme / Femme / Autre				
ET	Diagnostic(s)	Terminologie(s)	...	J45 asthme	Expl	Source de données	Type de diagnostic
ET	Analyse(s) biologique(s)	Type de l'analyse	...	Commencer à saisir			
ET	Sejour(s)	Date d'entrée dans l'épisc	=	2018-04-30			

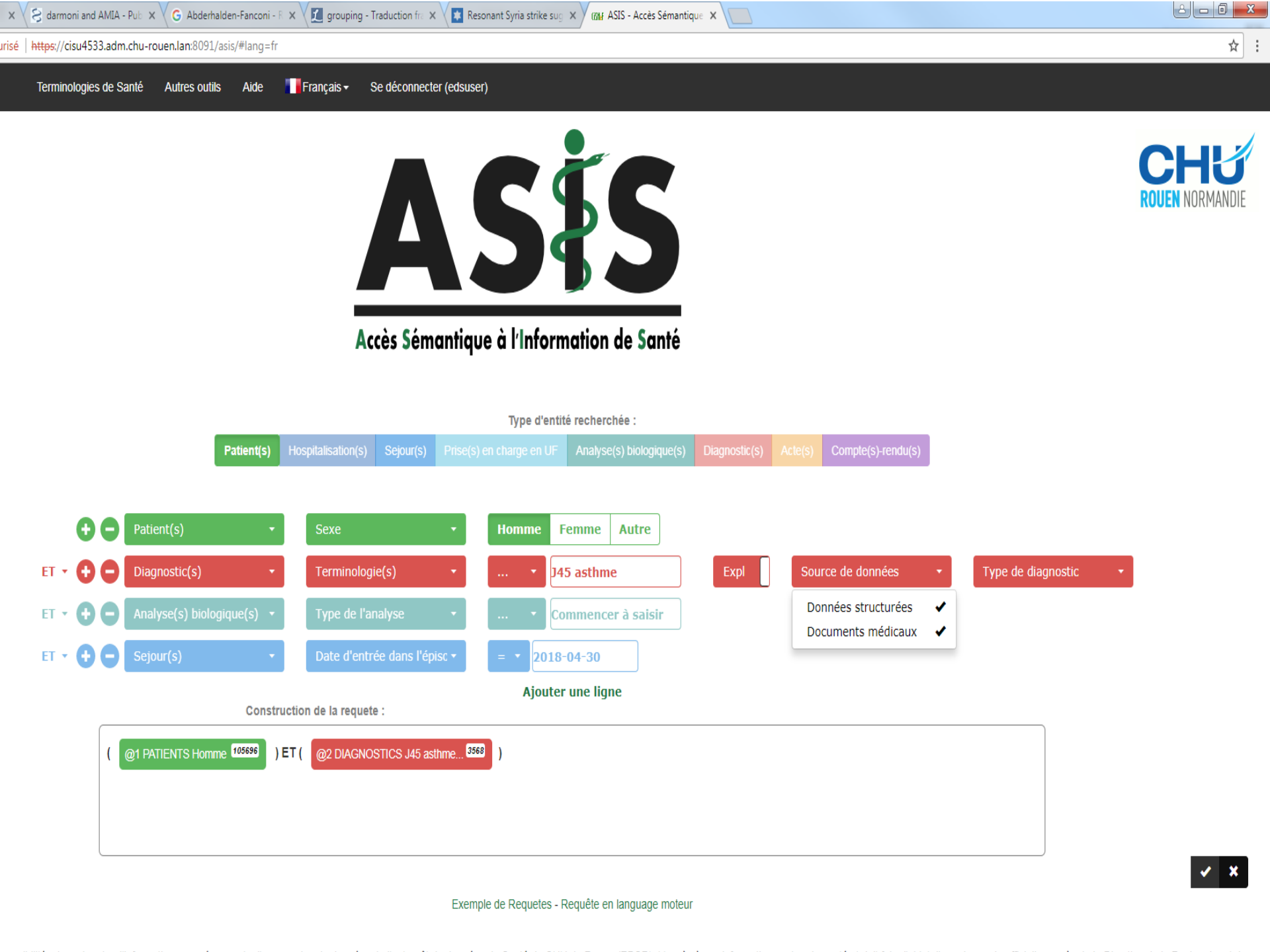
Source de données : Données structurées ✔ / Documents médicaux

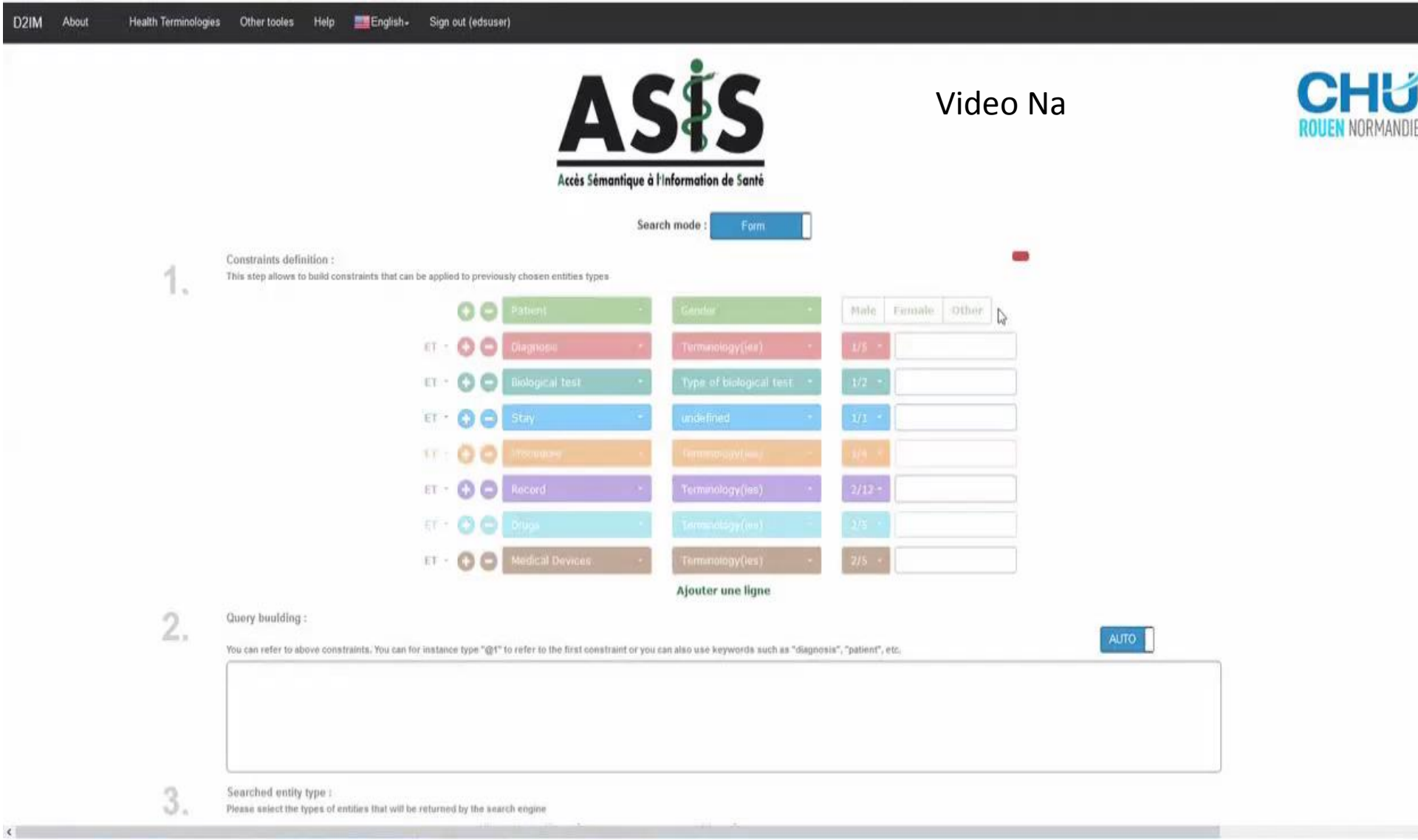
Ajouter une ligne

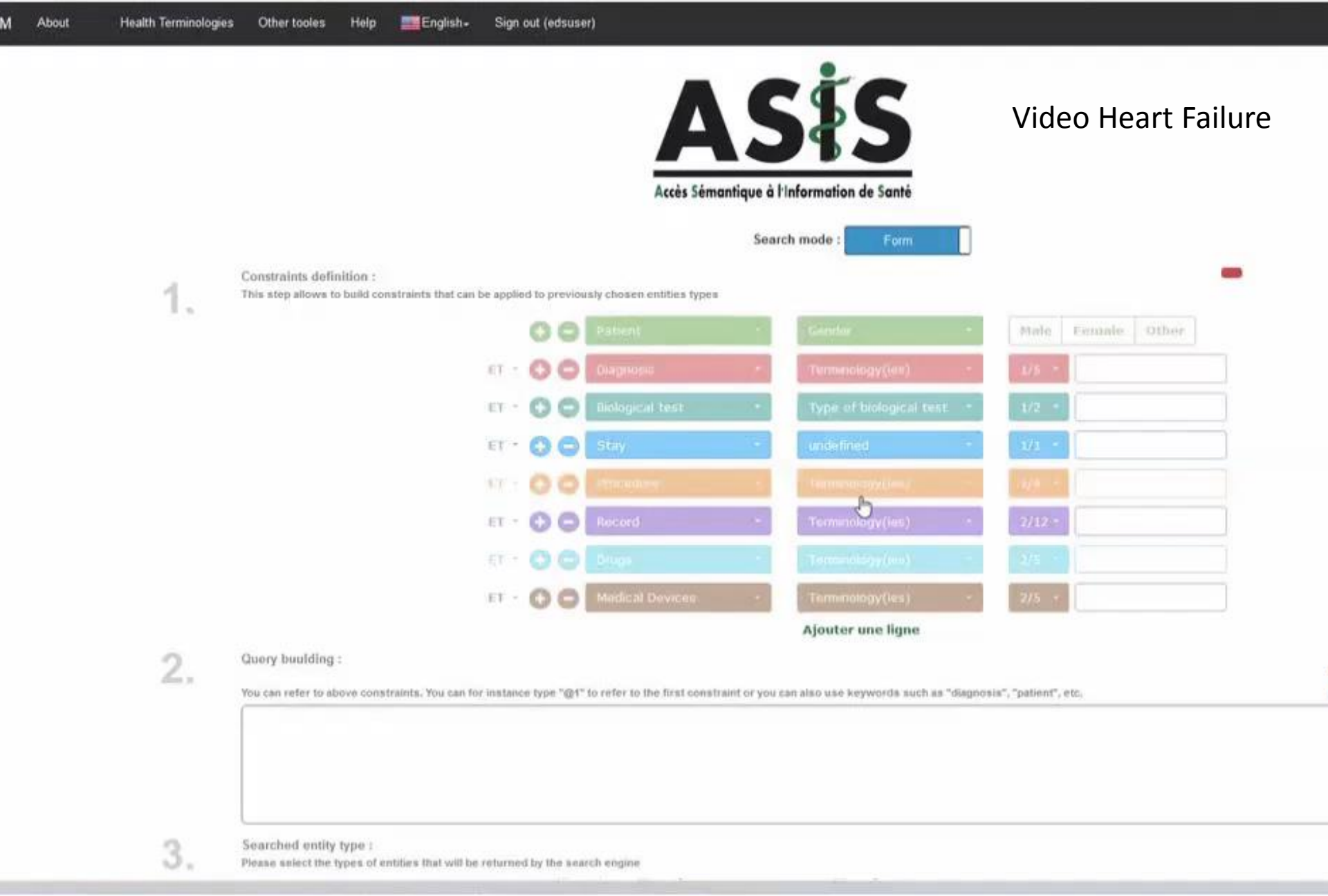
Construction de la requete :

(@1 PATIENTS Homme 105696) ET (@2 DIAGNOSTICS J45 asthme... 625)

Exemple de Requetes - Requête en language moteur


darmoni and AMIA - Pub
Abderhalden-Fanconi - F
grouping - Traduction fr
Resonant Syria strike sug
ASIS - Accès Sémantique
https://cisu4533.adm.chu-rouen.lan:8091/asis/#lang=fr
Terminologies de Santé
Autres outils
Aide
Français
Se déconnecter (edsuser)
ASIS
Accès Sémantique à l'Information de Santé
CHU ROUEN NORMANDIE
Type d'entité recherchée :
Patient(s)
Hospitalisation(s)
Sejour(s)
Prise(s) en charge en UF
Analyse(s) biologique(s)
Diagnostic(s)
Acte(s)
Compte(s)-rendu(s)
Patient(s)
Sexe
Homme
Femme
Autre
ET
Diagnostic(s)
Terminologie(s)
...
J45 asthme
Expl
Source de données
Type de diagnostic
Données structurées
Documents médicaux
ET
Analyse(s) biologique(s)
Type de l'analyse
...
Commencer à saisir
ET
Sejour(s)
Date d'entrée dans l'épisc
=
2018-04-30
Ajouter une ligne
Construction de la requete :
(@1 PATIENTS Homme 105696) ET (@2 DIAGNOSTICS J45 asthme... 3568)
Exemple de Requetes - Requête en language moteur


D2IM
About
Health Terminologies
Other tooles
Help
English
Sign out (edsuser)
ASIS
Accès Sémantique à l'Information de Santé
Video Na
CHU
ROUEN NORMANDIE
Search mode :
Form
1.
Constraints definition :
This step allows to build constraints that can be applied to previously chosen entities types
Patient
Gender
Male
Female
Other
ET
Diagnosis
Terminology(ies)
1/5
Biological test
Type of biological test
1/2
Stay
undefined
1/1
Record
Terminology(ies)
2/12
Drugs
2/5
Medical Devices
Terminology(ies)
2/5
Ajouter une ligne
2.
Query buulding :
You can refer to above constraints. You can for instance type "@1" to refer to the first constraint or you can also use keywords such as "diagnosis", "patient", etc.
AUTO
3.
Searched entity type :
Please select the types of entities that will be returned by the search engine



Video Heart Failure

Valorization

- **Alicante**
  - SME from the North of France
  - Exclusive sales of RUH DBI tools: HeTOP & ECMT => two French hospitals
  - Integration of the semantic search engine in 2019

- **OMICX**
  - One PhD to develop deep learning tools for annotation & information retrieval (sept. 2017 + sept. 2018)
  - Licence fee for semantic search engine (2017-8)

Use cases of the sHDW

- Since March 2018, over 60 use cases treated from SHDW (still POC)
  - Huge feedback to modify our top priorities to be developed
  - One example: to heavily invest on the semantic annotator as 80% of the questions were directly linked to health documents vs. lab tests or drugs

- One example: all lab tests during several years for specific infants

- In June 2018, to identify all the patients with endocarditis AFTER TAVI Transcatheter Aortic Valve Implantation
- To identify the TAVI acronym in the documents, thanks to the semantic annotator (ECMT)
  - Number of cases retrieved based on DRGs data warehouse = 30
  - Number of cases retrieved based on current RUH data warehouse = 53 (2 false negatives)
  - 23 new cases retrieved thanks to EDSaN!!!

And soon…

- « v1 » to be launched (2019-2020)
  - Industrialization (Alicante)
  - « generic & unified » patient centric model +++
  - More data (drugs, medical devices, resuscitation, etc.)
  - Management of temporal data
    - Allen algebra (BEFORE, DURING, AFTER)
  - Query library
  - Export module for DRG department (SQL)

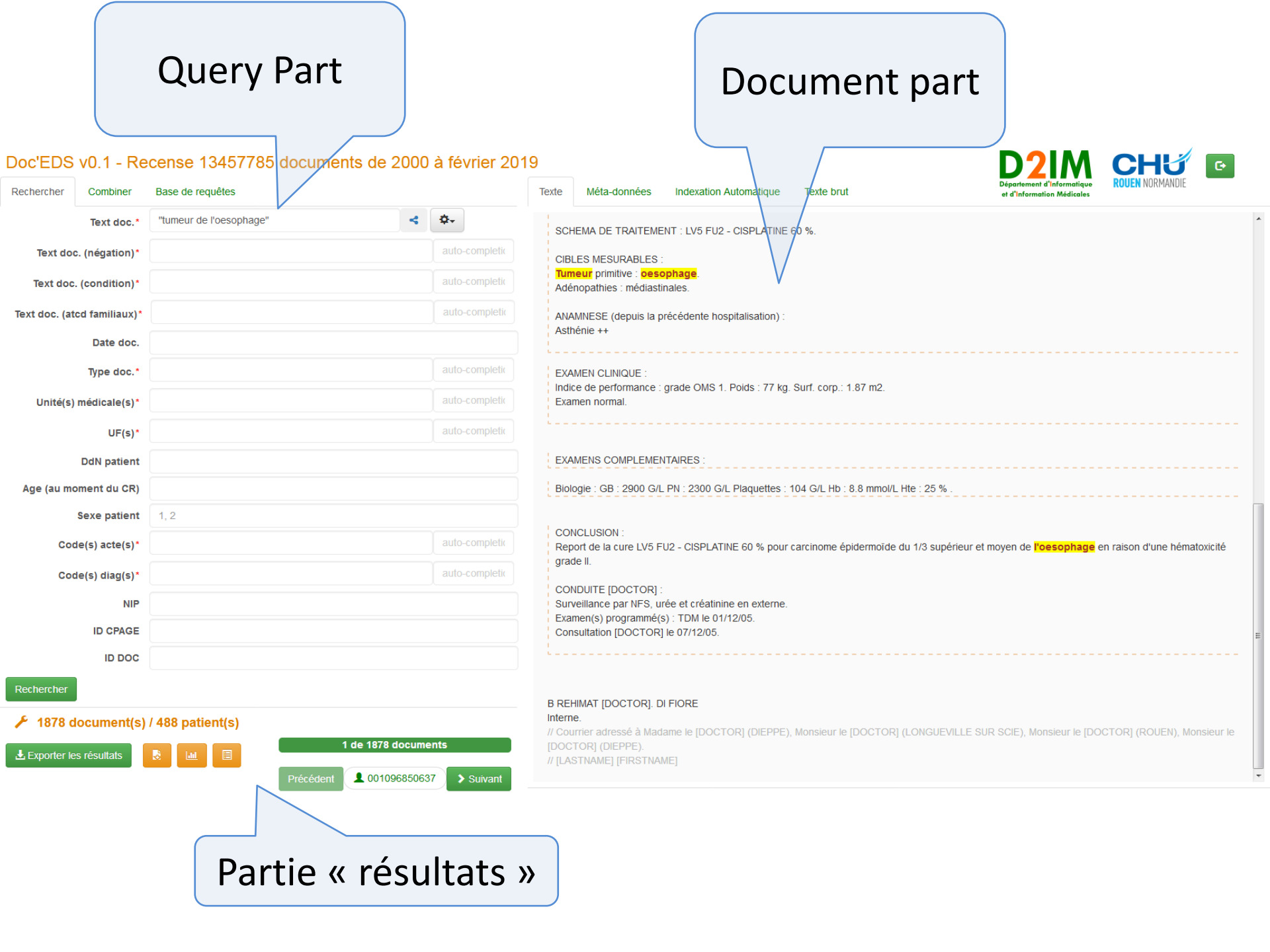
- HDW not limited to RUH: data from other hospitals, private offices, national data
- Health Data Hub in France ≠ Israeli (or Swiss) HDW

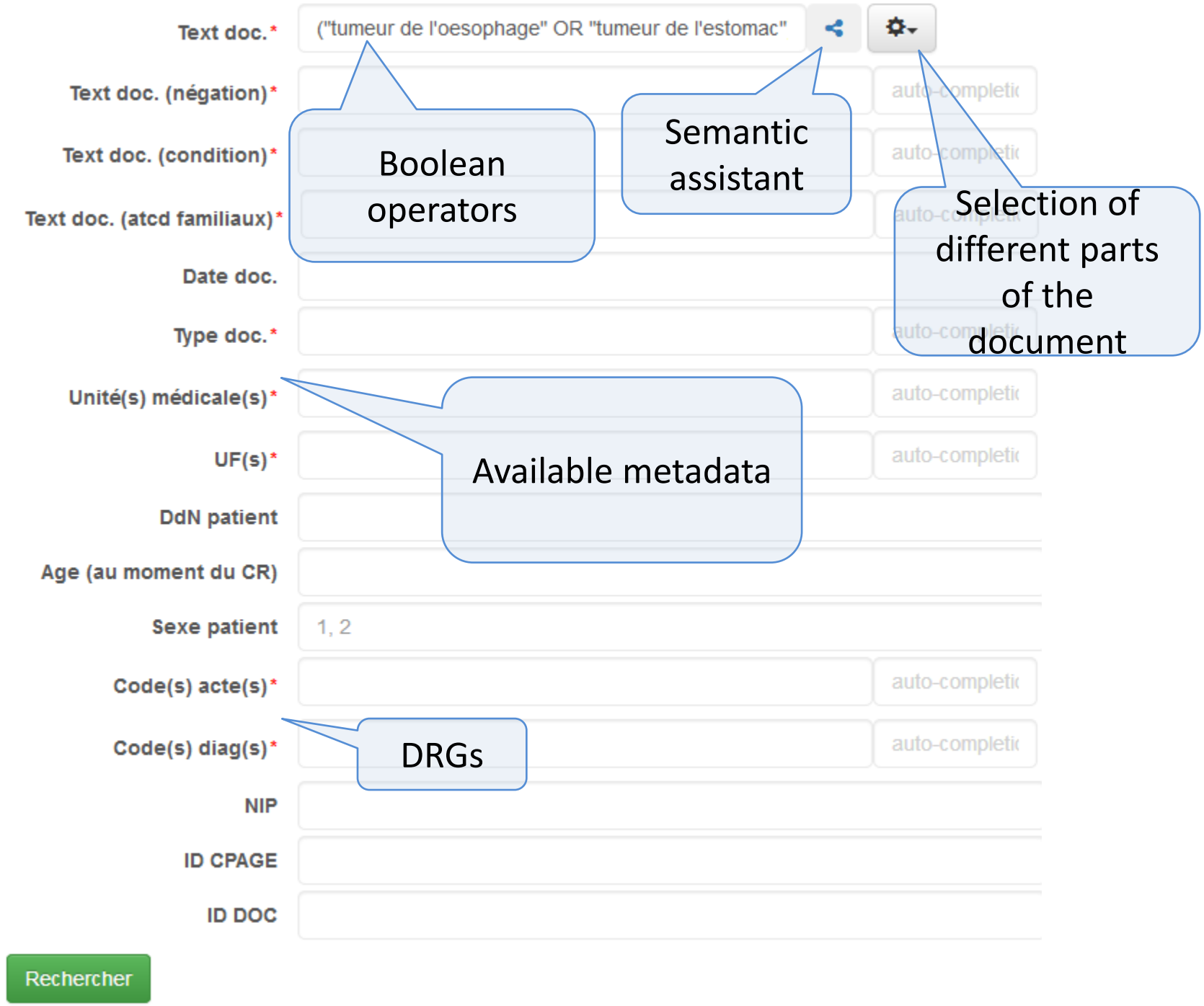
Doc'EDS

- New tool developed in 2019, based on the first 20 use cases since March 2018
- No semantics… for now => better cost-effectiveness
- Based on Lucene
- To better understand the deeper structure of the document (80% of the health information in France)
  - Negation
  - Hypothesis
  - Family history
  - Segmentation of the document

Doc'EDS

- Allow « manual dataminig »: oxymoron!
- Tool to validate the quality of the queries: continuous improvment of the search engine
- Based on rules, regular expressions and frequency analyses


Query Part
Document part
Doc'EDS v0.1 - Recense 13457785 documents de 2000 à février 2019
Rechercher
Combiner
Base de requêtes
Text doc.*
"tumeur de l'oesophage"
Text doc. (négation)*
Text doc. (condition)*
Text doc. (atcd familiaux)*
Date doc.
Type doc.*
Unité(s) médicale(s)*
UF(s)*
DdN patient
Age (au moment du CR)
Sexe patient
1, 2
Code(s) acte(s)*
Code(s) diag(s)*
NIP
ID CPAGE
ID DOC
Rechercher
1878 document(s) / 488 patient(s)
Exporter les résultats
1 de 1878 documents
Précédent
001096850637
Suivant
Texte
Méta-données
Indexation Automatique
Texte brut
D2IM
Département d'Informatique et d'Information Médicales
CHU ROUEN NORMANDIE
SCHEMA DE TRAITEMENT : LV5 FU2 - CISPLATINE 60 %.
CIBLES MESURABLES :
Tumeur primitive : oesophage.
Adénopathies : médiastinales.
ANAMNESE (depuis la précédente hospitalisation) :
Asthénie ++
EXAMEN CLINIQUE :
Indice de performance : grade OMS 1. Poids : 77 kg. Surf. corp.: 1.87 m2.
Examen normal.
EXAMENS COMPLEMENTAIRES :
Biologie : GB : 2900 G/L PN : 2300 G/L Plaquettes : 104 G/L Hb : 8.8 mmol/L Hte : 25 % .
CONCLUSION :
Report de la cure LV5 FU2 - CISPLATINE 60 % pour carcinome épidermoïde du 1/3 supérieur et moyen de l'oesophage en raison d'une hématoxicité grade II.
CONDUITE [DOCTOR] :
Surveillance par NFS, urée et créatinine en externe.
Examen(s) programmé(s) : TDM le 01/12/05.
Consultation [DOCTOR] le 07/12/05.
B REHIMAT [DOCTOR]. DI FIORE
Interne.
// Courrier adressé à Madame le [DOCTOR] (DIEPPE), Monsieur le [DOCTOR] (LONGUEVILLE SUR SCIE), Monsieur le [DOCTOR] (ROUEN), Monsieur le [DOCTOR] (DIEPPE).
// [LASTNAME] [FIRSTNAME]
Partie « résultats »


Text doc.*
("tumeur de l'oesophage" OR "tumeur de l'estomac"
Text doc. (négation)*
Text doc. (condition)*
Text doc. (atcd familiaux)*
Date doc.
Type doc.*
Unité(s) médicale(s)*
UF(s)*
DdN patient
Age (au moment du CR)
Sexe patient
1, 2
Code(s) acte(s)*
Code(s) diag(s)*
NIP
ID CPAGE
ID DOC
Rechercher
Boolean operators
Semantic assistant
Selection of different parts of the document
Available metadata
DRGs

RAMI[DOCTOR] 5 mg, un par jour
FUROSEMIDE 20 mg, un par jour
SELOKEN 200 mg, 1 par jour
SINVASTATINE 40 mg, un par jour
INNOHEP 14000, une injection par jour
IMOVANE 7.5 mg, un au coucher
VESICARE 5 mg, un par jour
UROREC 8 mg, un par jour

Anonymization (patients and physicians)

MODE DE VIE
Marié (femme aidante)

HISTOIRE DE LA MALADIE :
Patient de [redacted] ans hospitalisé le matin [redacted]/04 à Becquerel pour une cure de radiothérapie dans le cadre de la prise en charge d'un carcinome épidermoide de **l'oesophage** suivit au CHU en gastro-entérologie (médecin référent : [DOCTOR]).
Le patient décrit des expectorations mêlées de sang survenues en deuxième moitié de nuit (nuit du [redacted] au [redacted]/04). Un second épisode de crachats sanglants après effort de toux au cours de l'évaluation pré radiothérapie. Dans le contexte de **tumeur** de **l'oesophage**, le patient est transféré en unité de soins intensifs de gastro-entérologie au CHU pour une probable hématémèse.

Detection of suspiscion/hypotheses/doubts/future

EXAMEN CLINIQUE :
Taille : 176 cm : 76 kg. IMC : 24.5 kg/m². Température : 36.4°C Pouls : 77 par mn. TA : 109/70. *EVA : 0.* Hb : 12.6 g/Dl. Bonne tolérance hémodynamique de l'hémorragie, *pas de déglobulisation*
Dénutrition. OMS 2-3.
Conjonctives colorées. *Absence d'ictère.*
Cardiovasculaire : BDC irrégulières. *Pas de souffle. Pas d'OMI Pas de Turgescence des jugulaires.*
Respiratoire : Râles crépitants à la base pulmonaire gauche. Toux importantes suivies d'expectorations teintées de sang.
Digestif : présence d'une sonde naso-gastrique d'alimentation entérale. Abdomen souple, indolore. BHA présents et normaux. *Absence d'argument en faveur d'une ascite clinique.* Toucher rectal : selles dures de couleur normale (pas de méléna *ni de rectorragies*). Prostate de taille augmentée avec disparition du sillon médian.
Neurologique : bonne orientation temporo-spatiale. *Pas de déficit sensitivo-moteur.*

Détection of negations

BIOLOGIE :
Hémoglobine à 12,3 g/dl. .Sérologie et antigénurie aspergillaire en cours.

RADIOLOGIE :
Radio de thorax :
TDM du [redacted]02/20[redacted] : Stabilité de l'épaississement circonférentiel oesophagien avec multiples adénomégalies médiastinales.Apparition de plusieurs nodules lobaires inférieures gauches :

SCHEMA DE TRAITEMENT : Cacis (Campto + Cisplatine) + radiothérapie.

CIBLES MESURABLES :
Tumeur primitive : oesophage.
Adénopathies : oui.

ANAMNESE (depuis la précédente hospitalisation) :

Antécédents : chirurgie sur le ménisque droit. Chirurgie sur varices superficielles des membres inférieurs. Terrain atopique : eczéma.
Frère décédé d'un cancer du poumon il y a 30 ans.
Oncle décédé d'un cancer de l'oesophage il y a 16 ans.

EXAMEN CLINIQUE :
Indice de performance : grade OMS 0. Poids : 89 kg. Surf. corp.: 2 m2.

EXAMENS COMPLEMENTAIRES :

Biologie : GB : 6600 G/L PN : 3890 G/L Plaquettes : 272000 G/L Hb : 13,5 g/dl. TP : 99 %.
Urée : 0,24. Créat : 7,3 mg/l. TCA : 1,18.

TRAITEMENT RECU :
Chimiothérapie :
- Campto 60 mg/m2/j soit 120 mg/j (soit 100 % de la dose standard).
- Cisplatine 30 mg/m2/j soit 60 mg/j (soit 100 % de la dose standard).
- Traitement anti-émétique préventif par Zophren.
- Traitement anti-émétique curatif par Solumédrol.

Radiothérapie : à prévoir.

« Segmentation » du document

Détection des antécédents familiaux

Analyses

Différents bilans

Statistiques & bilans

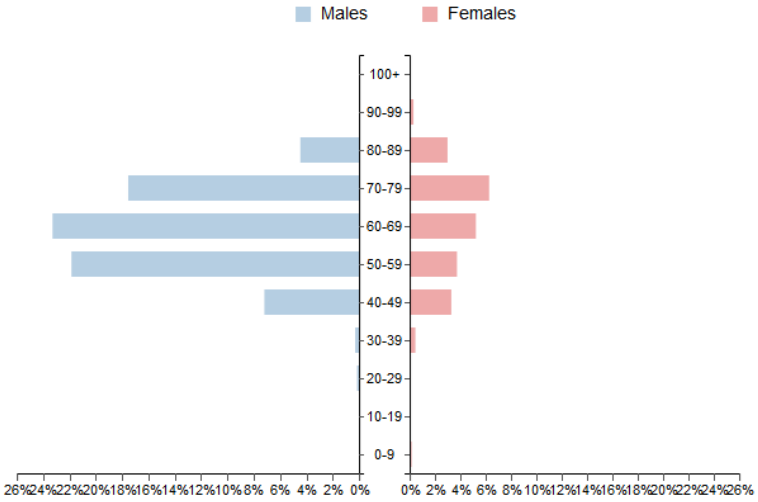
- Stats démographiques
- Date doc.
- Type doc.
- Unité(s) médicale(s)
- UF(s)
- DdN patient
- Age (au moment du CR)
- Sexe patient
- Code(s) acte(s)
- Code(s) diag(s)
- 

150 femmes (22,4%)

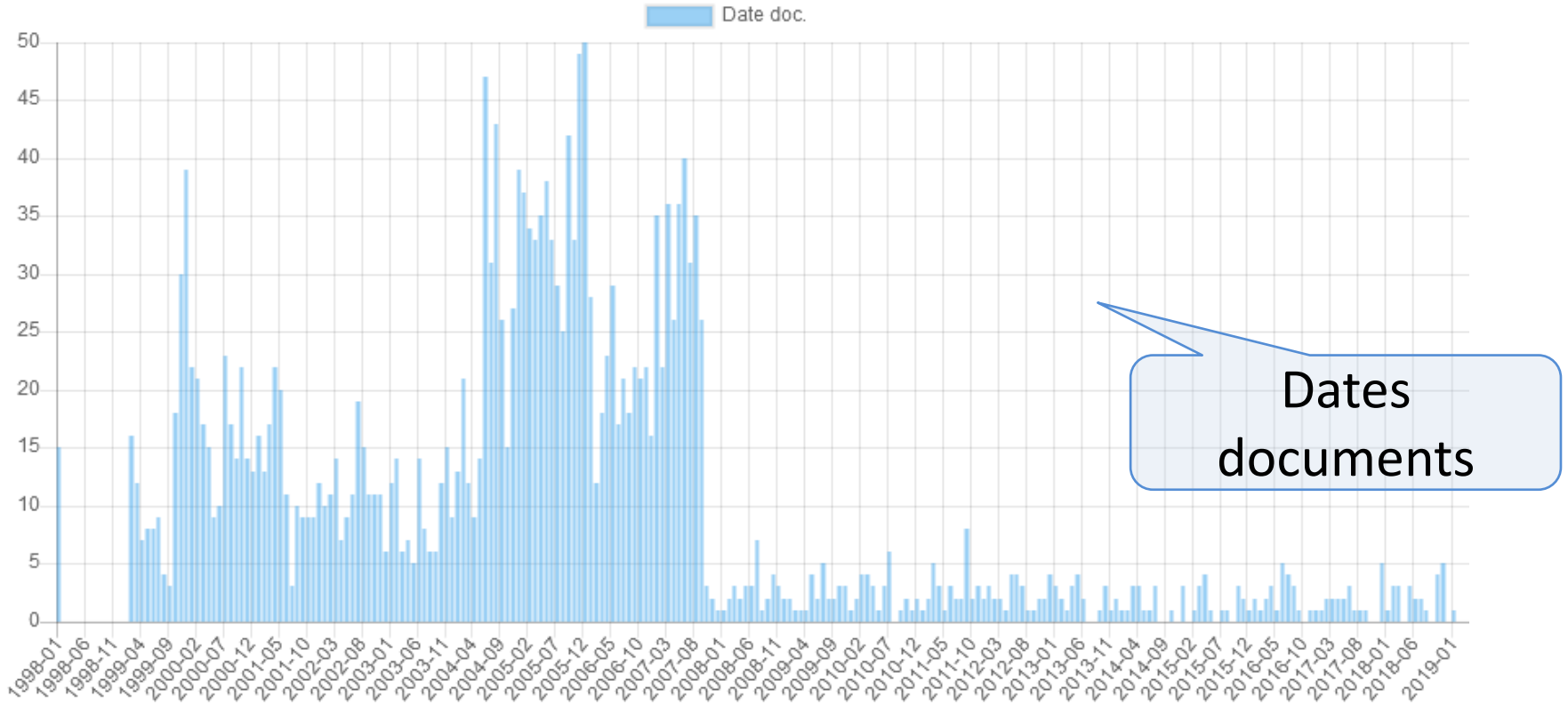
520 hommes (77,6%)

Statistiques des âges (au moment du CR)

Moyenne	63,3
Écart type	11,5
Minimum	7
Maximum	93
Q1	54
Q2	63
Q3	72



Population



SEJUM	#	%
DIGE HEPATO GASTRO ENTEROLOGIE NUTRITION	1685	61,9%
CGCD CHIRURGIE GENERALE ET DIGESTIVE	235	8,6%
RADI IMAGERIE CENTRALE	149	5,5%
URGE URGENCES	128	4,7%
PHIE PHARMACIE	93	3,4%
ORLO O.R.L ADULTES	61	2,2%
PNM1 CLINIQUE PNEUMOLOGIQUE HCN	59	2,2%
PHYS PHYSIOLOGIE DIGESTIVE	34	1,2%
CARD CARDIOLOGIE	25	0,9%
REAC REANIMATION CHIRURGICALE	23	0,8%

Répartition dans les unités médicales / services

Diagnostics
PMSI, etc.

Code(s) diag(s) | DP | DA | DR

Afficher 10 **éléments** **Filtrer :**

155 valeurs distinctes, 2369 occurrences au total

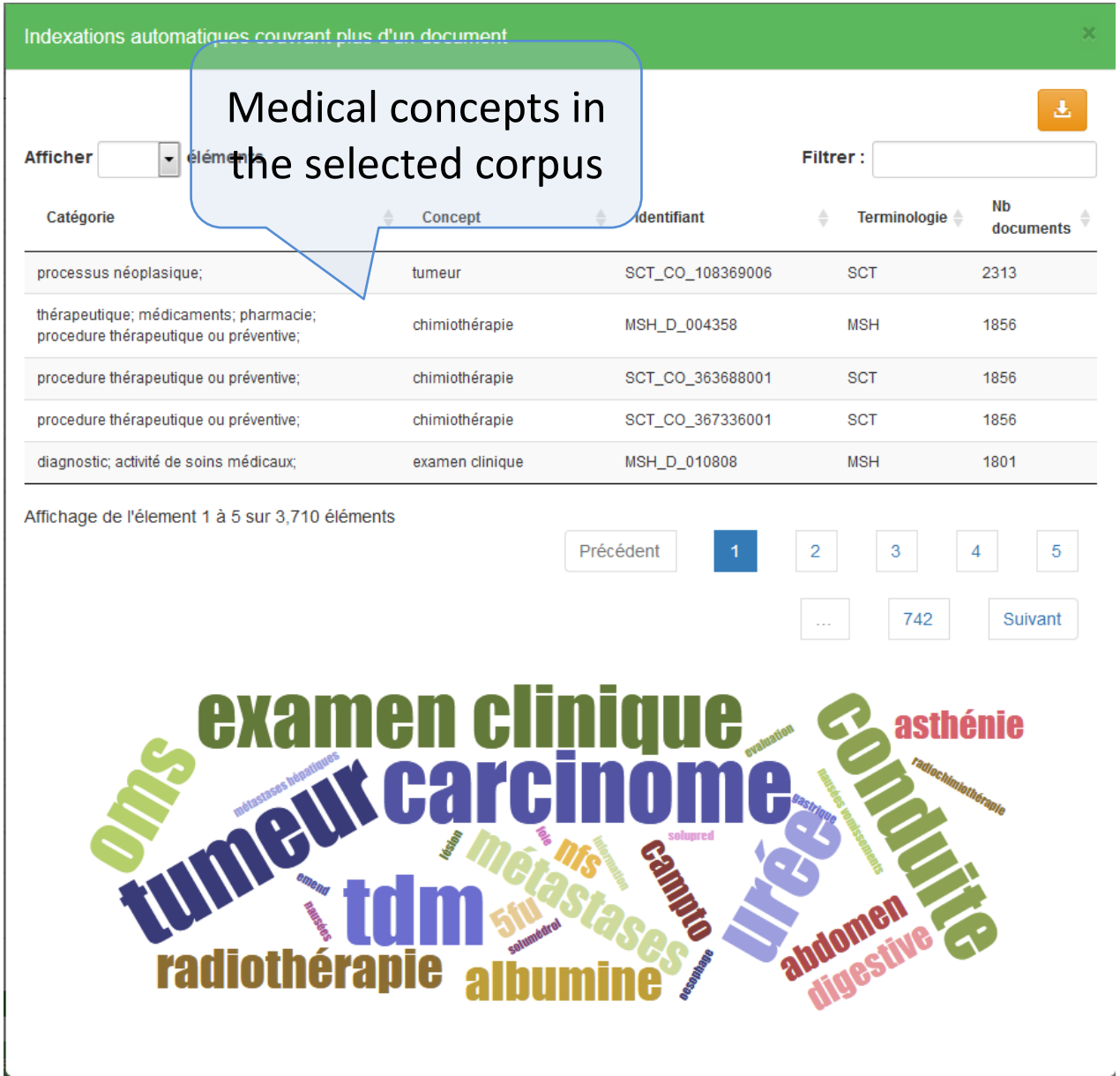
DIAGCODES	#	%
Z511 séance de chimiothérapie pour tumeur	1292	54,5%
Z530 acte non effectué en raison de contre-indication	302	12,7%
C155 tumeur maligne du tiers inférieur de l'oesophage	86	3,6%
C151 tumeur maligne de l'oesophage thoracique	66	2,8%
Z087 examen de contrôle après traitements combinés pour tumeur maligne	63	2,7%
C154 tumeur maligne du tiers moyen de l'oesophage	50	2,1%
C150 tumeur maligne de l'oesophage cervical	39	1,6%
C153 tumeur maligne du tiers supérieur de l'oesophage	39	1,6%
Z452 ajustement et entretien d'un dispositif d'accès vasculaire	34	1,4%
C159 tumeur maligne de l'oesophage, sans précision	24	1,0%

Affichage de l'élement 1 à 10 sur 155 éléments

Précédent 1 2 3 4 5 … 16

Suivant

Data mining (ECMT)



Filtering (diseases, procedures, …) : here drugs

Afficher éléments

Filtrer : PHA_RAC

Catégorie	Concept	Identifiant	Terminologie	Nb documents
Substance pharmacologique; produit chimique inorganique;	cisplatine	PHA_RAC_1826	PHA	1150
chimie organique; Substance pharmacologique;	zophren	PHA_RAC_8895	PHA	1015
chimie organique; Substance pharmacologique;	campto	PHA_RAC_1470	PHA	517
chimie organique; Substance pharmacologique;	solumédrol	PHA_RAC_7444	PHA	401
médicament clinique;	solupred	PHA_RAC_7445	PHA	306

Affichage de l'élement 1 à 5 sur 332 éléments (filtré de 3,710 éléments au total)

Précédent 1 2 3 4 5 … 67 Suivant

Perspectives

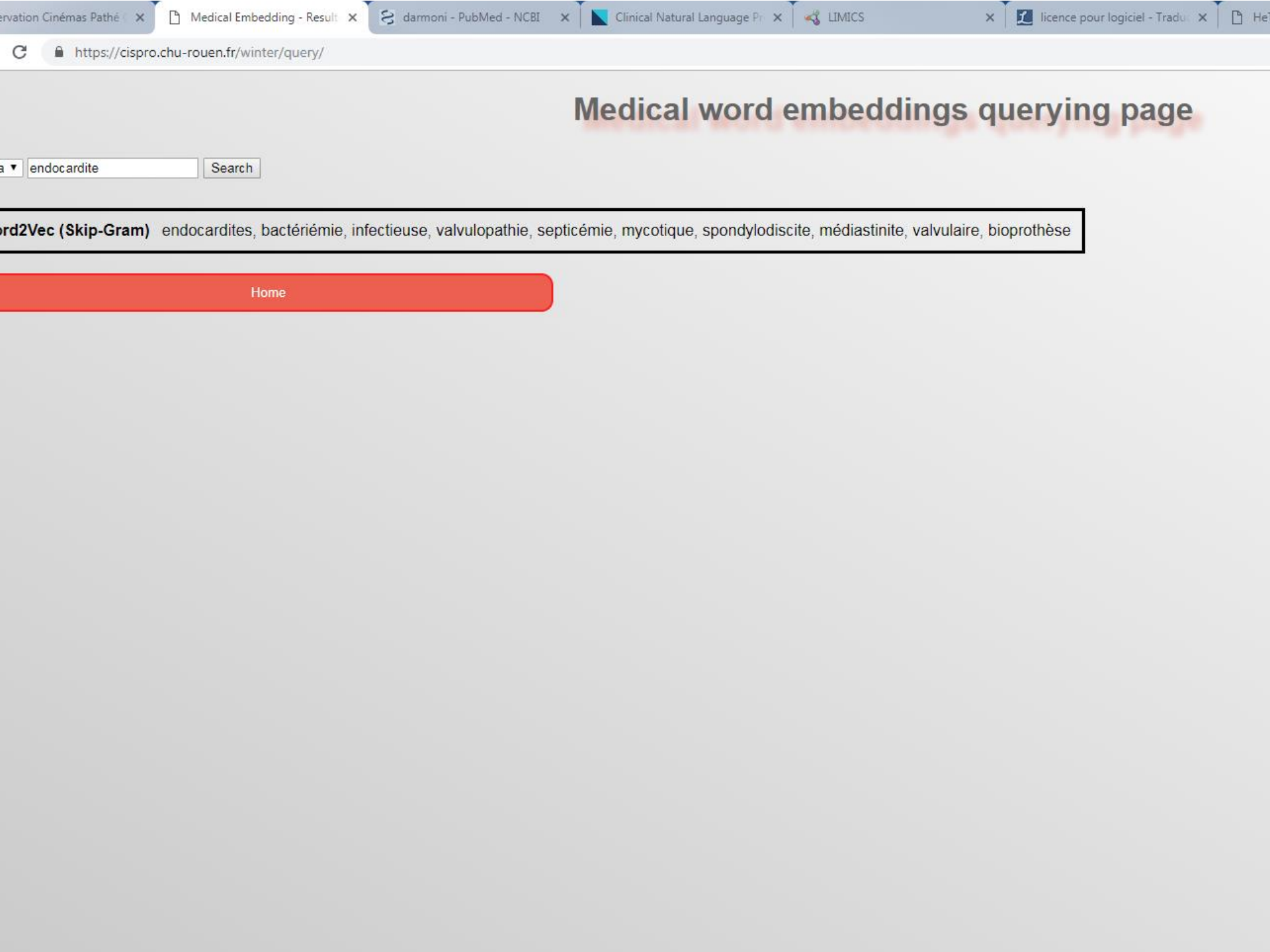
- Medical word embeddings (Word2Vec, FastText,Glove)
  - PhD Emeric Dynomant (OMICX)
  - Two different corpus
    - 12 M health documents from HSDW
    - 180 K abstracts from LiSSa, French bibliographic database
- Doc2Vec, Patient2Vec (in progress)
  - PhD Mikaël Dusenne, MD
    - Hybrid semantic annotator (“old” NLP + deep NLP)
    - Doc2Vec2DRGs, using an other tool !!! ELMO?

Medical word embeddings querying page

12M ▾ endocardite Search

GloVe	infectieuse, myocardite, eto, native, streptocoque, bovis, bactériémie, faecalis, _endocardite
FastText (CBOW)	proprio_septive, myopericardite, septo_optique, endo_aortique, endocartite, endoculaire, acrodermite, rhino_septale, salmonellose, épidermolyse
FastText (Skip-Gram)	endocartite, endocardique, proctologique, extancilline, septo_basale, prolongements, recanalisée, précentrale, dantrolene, podoscopique
Word2Vec (Skip-Gram)	bovis, sanguinis, eto, gordonii, gallolyticus, aorto_mitrale, mutans, infectieuse, streptoccoque, salivarius
Word2Vec (CBOW)	endocartite, _endocardite, native, bovis, médiastinite, myocardite, mutans, gallolyticus, myopéricardite, tamponnade

Home

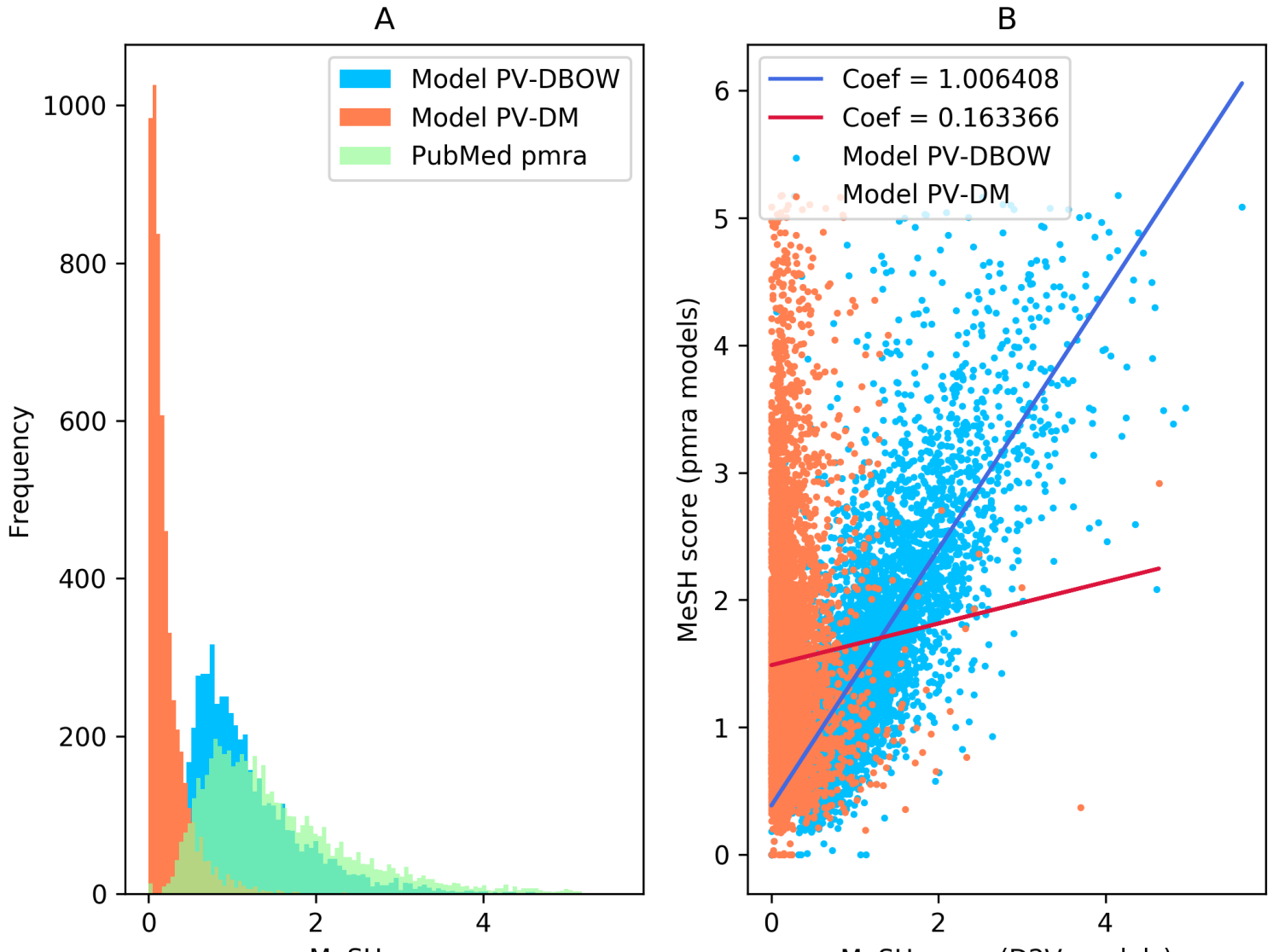

ervation Cinémas Pathé
Medical Embedding - Result
darmoni - PubMed - NCBI
Clinical Natural Language Pr
LIMICS
licence pour logiciel - Tradu
https://cispro.chu-rouen.fr/winter/query/
Medical word embeddings querying page
endocardite
Search
ord2Vec (Skip-Gram) endocardites, bactériémie, infectieuse, valvulopathie, septicémie, mycotique, spondylodiscite, médiastinite, valvulaire, bioprothèse
Home

Wordembeddings in two different contexts

QUERY: “facebook”

LiSSa corpus (300k)	internet, twitter, web, blog, e_learning, blogs, internautes, tic, game, …
RUH documents (12M)	reproches, injures, messages, insultes, rumeurs, ex_conjointe, menaces, insultant, …

Doc2Vec2PubMed



Thank you for listening!

Questions?

Email: Stefan.Darmoni@chu-rouen.fr