



# Doc'EDS



**Julien Grosjean, Badisse Dahamna, Romain Lelong, Ivan Kergourlay, Kévin Billey  
Thibaut Pressat-Laffouilhère, Pierre Balayé  
Jean-Philippe Leroy, Stéfan Darmoni**

*Département d'informatique et d'information médicale, CHU de Rouen*

*LIMICS U1142 INSERM*

**D2IM**  
Département d'Informatique  
et d'Information Médicales



# Un outil de recherche et de fouille des documents médicaux textuels

- Prendre en compte la structure du document et le discours (NLP) :
  - Négation
  - Condition/suspicion/hypothèse/futur
  - Antécédents familiaux
  - Segmentation du document
- Rapide
- Analyses macros par corpus

# Méthodes (1/2)

- Import des documents textuels (ou conversion si besoin types doc, docx, PDF, rtf...) éventuellement liés à des éléments de contexte (ID patient, ID séjour, codes PMSI, unité médicale, etc.)
1. Ignorer les entêtes et pieds de page (fréquence des lignes)
  2. Anonymisation (patient, famille, professionnels de santé, numéros, adresses, etc.)
  3. Segmentation (motifs/indication, antécédents, anamnèse, résultats, traitements, etc.)
  4. Négations
  5. Hypothèses diagnostic/suspensions, éléments futurs
  6. Antécédents familiaux

# Méthodes (2/2)

- Indexation Lucene
- Annotation ECMT optionnelle pour fonctions analytiques en aval
- Temps de traitement dépendants de la machine, pour CHU de Rouen, 16,5M documents sur un « gros serveur » (192 cœurs, 1To de RAM) :
  - Pour indexation Lucene : env. 4h
  - Pour ECMT : env. 22h
- Une fois les indexations faites, l'outil est ultra léger et ne dépend (presque) que des accès disque :
  - sur un Intel NUC i5 (400€) : temps réponse \* 1,07 vs « gros serveur »
  - sur un Raspberry Pie (120€) : temps réponse \* 4,58 vs « gros serveur »

# Partie « requête »

Doc'EDS Requête Sélection Cohorte Aide

Requête simple Historique/combiner

Texte "tumeur de l'oesophage"

Texte (atcd familiaux) décédé OR décès

Date doc. [2000-01-01,2001-01-01], <1999-01-01, >2018-05-15

Type doc. auto-complète

Unité(s) médicale(s) auto-complète

UF(s) auto-complète

DdN patient [1950-01-01,1960-01-01], <1930-01-01, >2010-06-03

Age (au moment du CR) >16

Sexe patient 1, 2

Code(s) acte(s) auto-complète

Code(s) diag(s) auto-complète

NIP

ID CPAGE

ID DOC

Rechercher

D2IM Département d'Informatique et d'Informations Médicales CHU ROUEN NORMANDIE

# Pat. 677 # Doc. 2417

Requête 1 de 2417 documents

Texte Méta-données Indexation Automatique Texte brut

CIBLES MESURABLES :  
Tumeur primitive : oesophage

TOLERANCE INTERCURE :  
- Nausées/vomissements : Grade OMS I.  
RAS par ailleurs.

ANAMNESE (depuis la précédente hospitalisation) :  
- Nausées grade I à J1 en post-cure, pas de vomissement.  
- Epistaxis intermittentes de faible abondance.  
- Pas de dysphagie aux solides.  
- Pas de toux à la déglutition, ni de fausse route

EXAMEN CLINIQUE :  
Indice de performance : grade OMS 0. Poids : 76 kg (- 1,5 kg). Surf. corp. : 1,85 m2.  
Anorexie.  
Evaluation calorique : 1400 kcal/jour (sans les compléments nutritionnels)

EXAMENS COMPLEMENTAIRES :  
Biologie : GB : 2,6 G/ [PERSON] : 1,92 G/L Plaquettes : 78 G/L Hb : 13,9 g/dl. Urée : 0,26. Créat : 9,7.  
NFS de contrôle : plaquettes : 80 000

TRAITEMENT RECU :  
Chimiothérapie :  
- Cure reportée en raison d'une thrombopénie grade I.  
- Autres traitements : Gaviscon - Fortimel 1/jour + 2 crèmes HP/jour

Partie « résultats »

Partie « document »

KAMI[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

#### 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 épidermoïde de l'oesophage suivi 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 hématomèse.

#### EXAMEN CLINIQUE :

Taille : 176 cm : 76 kg. IMC : 24.5 kg/m<sup>2</sup>. 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.

#### 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 :

Anonymisation  
(patients et  
médecins)

Détection des  
suspensions/hypothèses/doutes/futur

Détection des  
négations/infirmations

SCHEMA DE TRAITEMENT : Cacic (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

# 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)

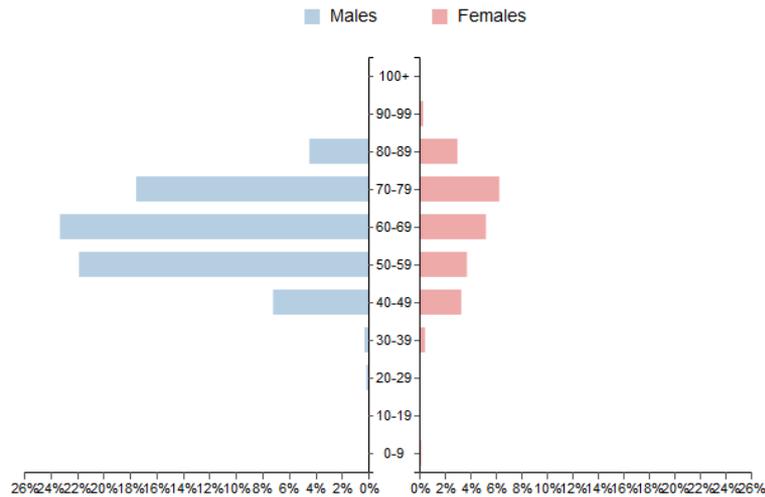


150 femmes (22,4%)

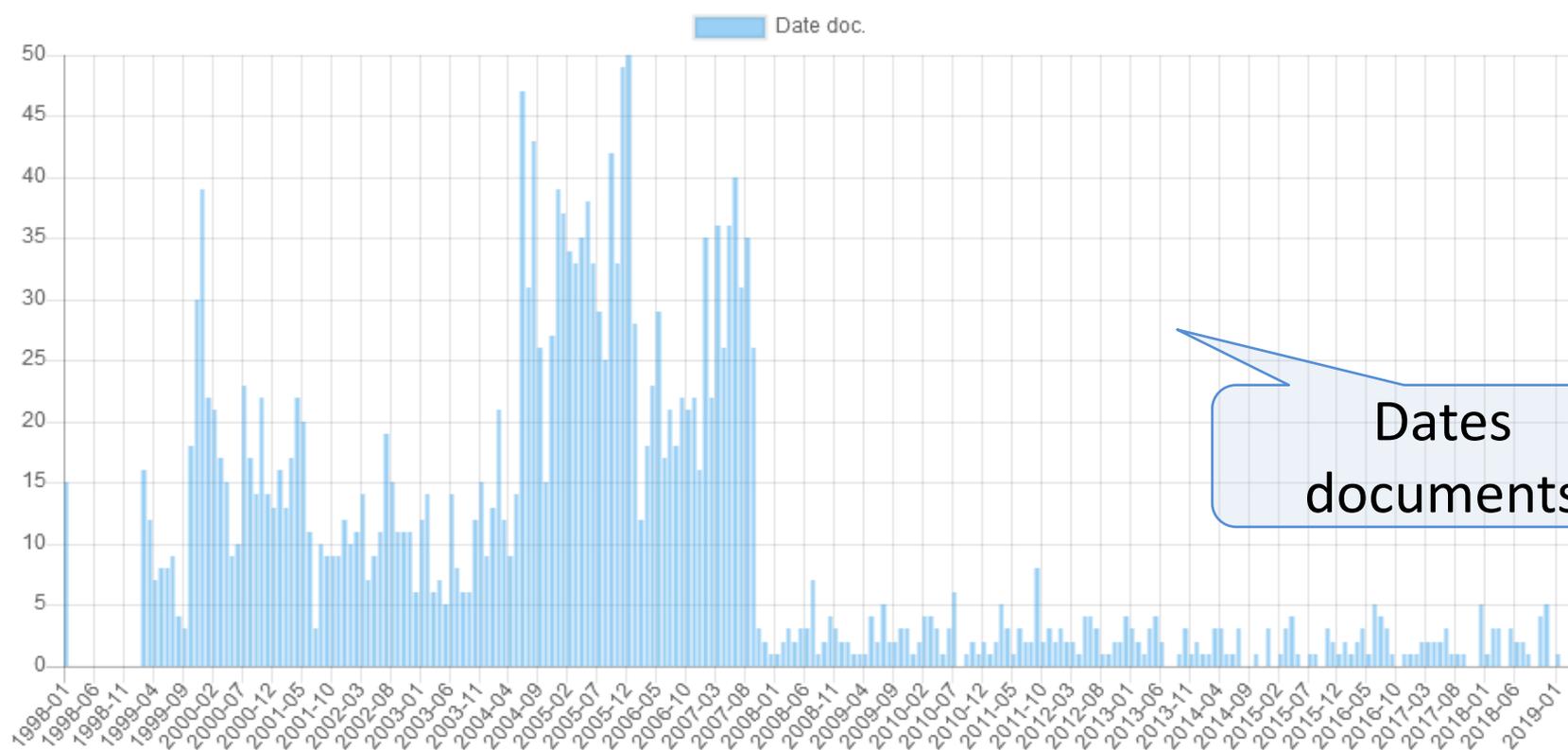
520 hommes (77,6%)

Statistiques des âges (au moment du CR)

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



Population



Dates documents

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'élément 1 à 10 sur 155 éléments

Précédent

1

2

3

4

5

...

16

Suivant

Les concepts médicaux présents dans le corpus



Afficher  éléments

Filtrer :

Catégorie	Concept	Identifiant	Terminologie	Nb documents
processus néoplasique;	tumeur	SCT_CO_108369006	SCT	2313
thérapeutique; médicaments; pharmacie; procédure thérapeutique ou préventive;	chimiothérapie	MSH_D_004358	MSH	1856
procédure thérapeutique ou préventive;	chimiothérapie	SCT_CO_363688001	SCT	1856
procédure thérapeutique ou préventive;	chimiothérapie	SCT_CO_367336001	SCT	1856
diagnostic; activité de soins médicaux;	examen clinique	MSH_D_010808	MSH	1801

Affichage de l'élément 1 à 5 sur 3,710 éléments



Filtrer (maladies, procédures, ...) : ici médicaments

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

Précédent

1

2

3

4

5

...

67

Suivant

# Démarche opérationnelle et Cas d'usage

# Démarche typique actuelle

- Comme pour construire des requêtes bibliographiques il y a des lustres...
- RDV avec un des « experts » Doc'EDS
- Présentation du sujet de recherche puis construction de la requête pas à pas
- Préparation de la méthode d'analyse des données (soit au sein de l'outil, soit en revenant sur les dossiers dans les services)
- Différents exports possibles

# Vers un circuit rationalisé

- Projet de plateforme web d'appui à la recherche unifiée avec la DRCI
- Demandes en ligne, support au travail collaboratif avec notamment le DIM et le DARC (bio- statistiques pour la méthodologie et les analyses)
- Travail et suivi sur la plateforme (dépôt de fichiers)
- Respect des normes RGPD

# Cas d'usage (N = 72)

- Cas « princeps » en juin 2018, liste anonyme de patients atteints d'endocardite après une pose de TAVI :
  - Travail en complément du DIM : identifier dans les documents textuels la mention de TAVI ; exploitation de l'annotateur automatique (ECMT)
  - Nombre de cas retrouvés par l'entrepôt DIM = 30
  - Nombre de cas retrouvés par l'entrepôt Rouen étendu = 53 (2 non retrouvés, présents dans l'entrepôt DIM)
  - 23 ont pu l'être grâce à l'exploitation des données non structurées présentes dans documents de santé (et traitement par l'ECMT)

# Cas d'usage Pharmacovigilance

- Prescriptions d'Androcur® (cyprotérone) à l'hôpital entre 2017 et 2019 (3 médecins à identifier pour communication des recommandations en vigueur)
- Questions éthiques... praticiens ET patients (*quid* des patients suivis ailleurs ?)

# Cas d'usage Lubrizol

- une centaine de cas dans les 10 premiers jours de passage aux urgences avec le terme « Lubrizol » dans les notes des urgentistes

# Bilan des cas d'usage tests

- 80% des cas d'usage pour mémoires/thèse internes médecine/pharma
- 95% des cas d'usage pour du pré-screening ; reste = pv, indicateurs de qualité, suivi cohorte
- Intérêt de l'outil pour consulter les dossiers hors spécialité
- Très bons retours dans l'ensemble : « gain de temps »
- Nos « meilleurs clients » : Dr Mehdi Taalba (Urgences), Dr Florence Tétart (Dermatologie) ; merci 😊

# Perspectives de l'outil

- Travail continu d'amélioration des algorithmes NLP
- Outil d'extraction semi-automatisée d'éléments textuels
- Déploiement dans des établissements de type/taille variés (outil « gratuit » en POC) ; projets de recherche +++

# Merci

[julien.grosjean@chu-rouen.fr](mailto:julien.grosjean@chu-rouen.fr)

Stefan.darmoni@chu-rouen.fr