

# Outils de recherche textuelle pour l'optimisation du codage PMSI

Journée du CRAIM – 14 novembre 2024 – N.Depas



- Développement des outils de recherche textuelle +/- IA sur les dernières années
  - Promesse d'un codage automatisé
  - Limites : hiérarchisation des diagnostics ++, négation difficile à prendre en compte, texte libre avec les fautes de frappe...
- Développement de différents outils :
  - Commerciaux : Sancare, Collective Thinking, Alicante...
    - Sancare en place au CHUGA : test en 2019, utilisé pour la clôture 2023 et poursuivi en 2024
  - Au sein d'entrepôts de données de Santé (ex : PREDIMED sur CHUGA)
  - Analyses internes au sein des DIM si accès aux bases du DPI

# SANCARE

- Solution clé en main qui se branche sur le DPI
- Serveur dédié : fourni par Sancare mais implanté et administré par le CHUGA
- Accès à l'ensemble des documents (courriers, transmissions ciblés, pancarte, biologie si configurée...)
- Intégration des données PMSI :
  - Fichier RSS groupé
  - Ou synchronisation directe avec le logiciel PMSI (CORA)

- CQ pré paramétrés de recherche textuelle (> 400)
- Ciblage large de dossiers mais nécessite un paramétrage fin par établissement pour éviter les faux positifs

## Recherches Sancare

### Certaines maladies infectieuses et parasitaires

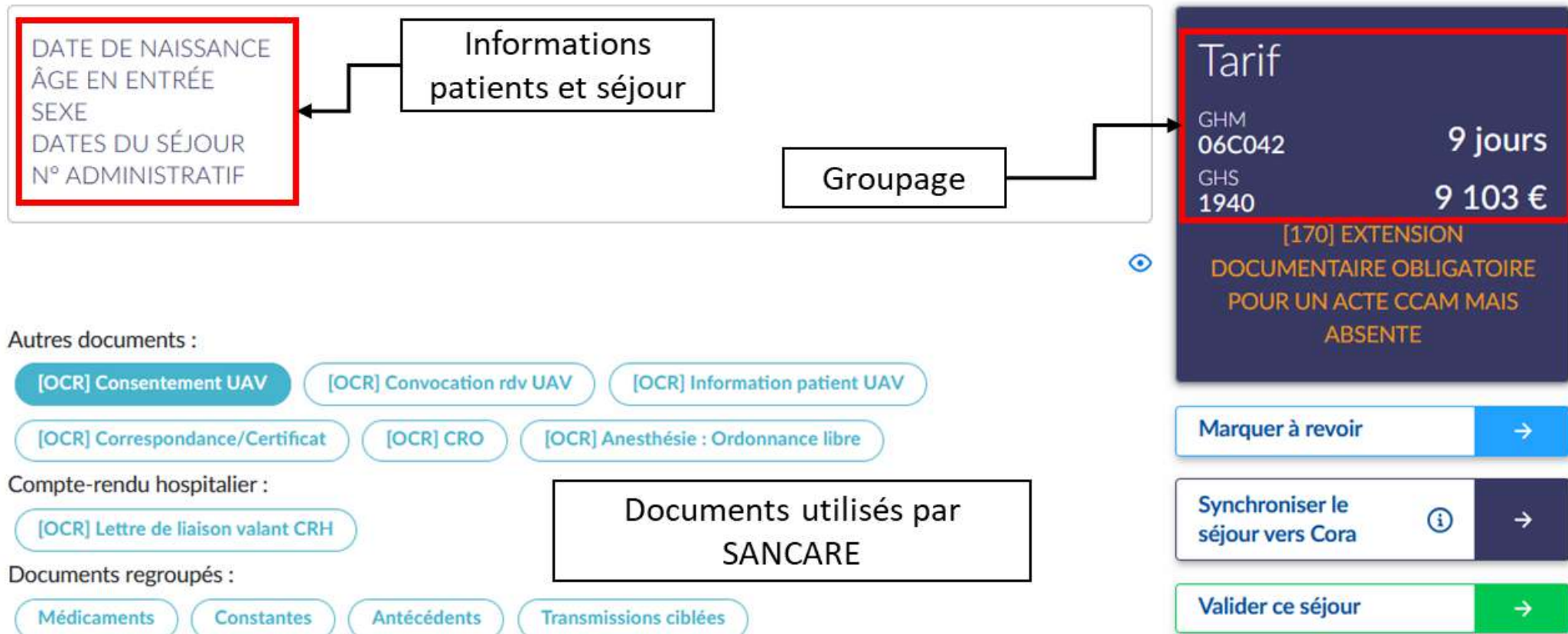
257 Voir tous les séjours de la catégorie	284 Staphylocoque Aureus	67 Entérocote à Clostridium difficile	57 Pneumocystose
53 Pseudomonas (pyocyanique)	52 Klebsiella pneumoniae	45 Sepsis à micro organisme anaérobie Gram négatif ou positif	25 Erysipèle
25 Tuberculose pulmonaire confirmée bactériologiquement	18 BLSE situation d'infection	18 Candidose, stomatite à candida	17 Tuberculose de l'appareil respiratoire sans confirmation bactéri
16 Sepsis à staphylocoques	15 Aspergillose pulmonaire invasive ou aspergillose disséminée	14 Candidose vulve et vagin	14 Mycoses superficielles
14 Streptocoques	11 E. Coli	10 Sepsis à Candida	7 Proteus Mirabilis ou Morganii
7 Staphylocoque	6 Autres salmonelloses sauf sepsis à Salmonella	5 Aspergillose pulmonaire, amygdalienne ou autres	5 Candidose (peau, ongles, phanères, seuls)
5 Zona (neuro, ophtalmo, disséminé)	4 Infections à Chlamydia transmises par voie sexuelle	4 Sepsis à micro-organismes anaérobies, Gram négatif ou positif av	4 Tuberculose de l'appareil respiratoire avec confirmation
2 Gale	2 Tuberculose des os et des articulations	1 Burkholderia Capacia	1 Candidose uro-génitale
1 Haemophilus Influenzae	1 Herpes ano-génital	1 Listériose sauf sepsis et méningite	1 Méningite virale
1 Pédiculose et phthiriasis	1 Sepsis à staphylocoques avec bio	1 Tuberculose du système nerveux	

### Maladies de l'appareil circulatoire

159 Voir tous les séjours de la catégorie	101 Hypertension pulmonaire	97 Endocardite Infectieuse	69 Hypotension idiopathique, orthostatique ou médicamenteuse
62 Arrêt cardiaque	57 Embolie pulmonaire	38 Hémorragie Intracérébrale grave S4	25 Infarctus cérébral AVC Ischémique, sans précision S2

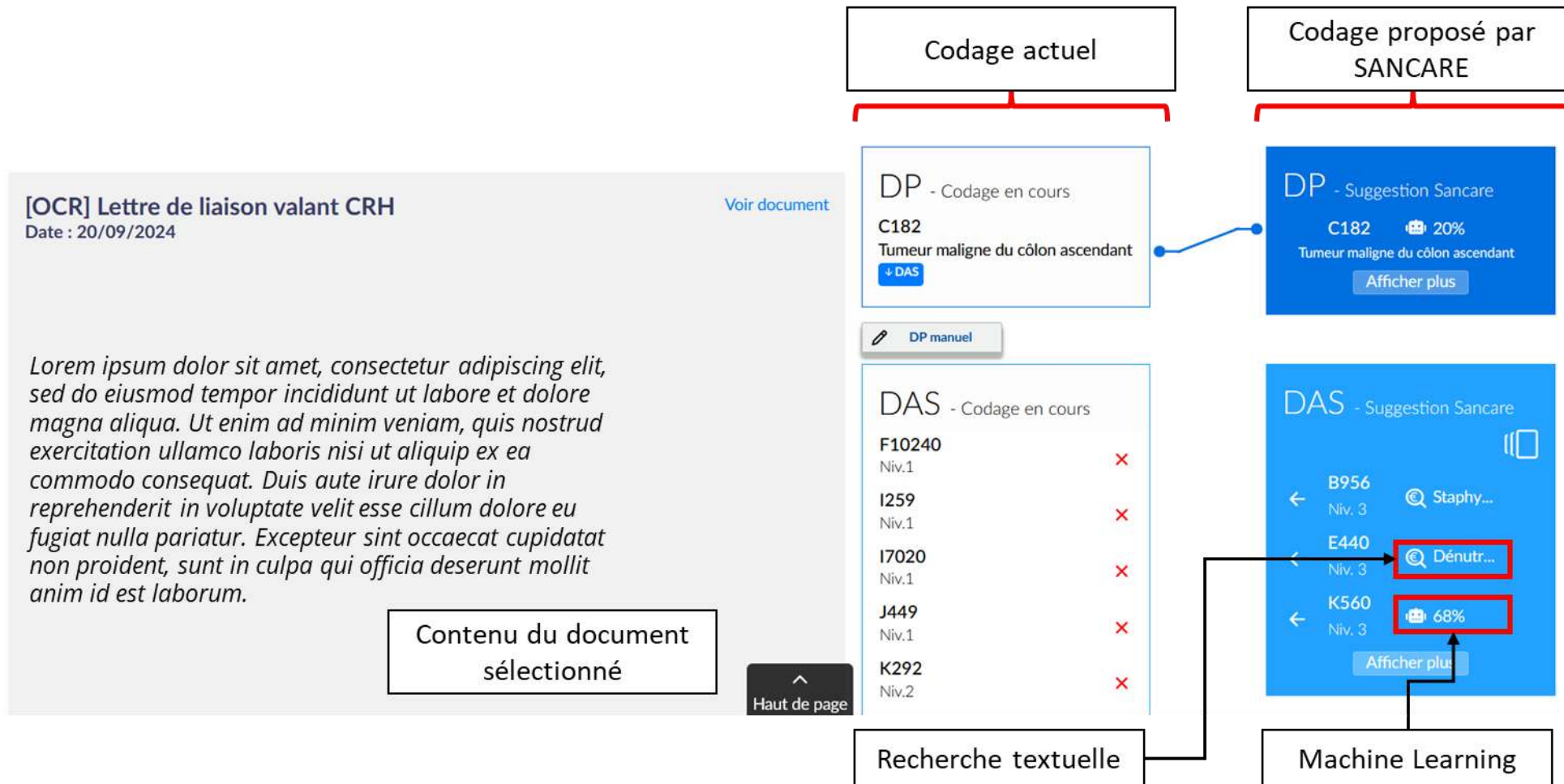


- 1 page par dossier revalorisable

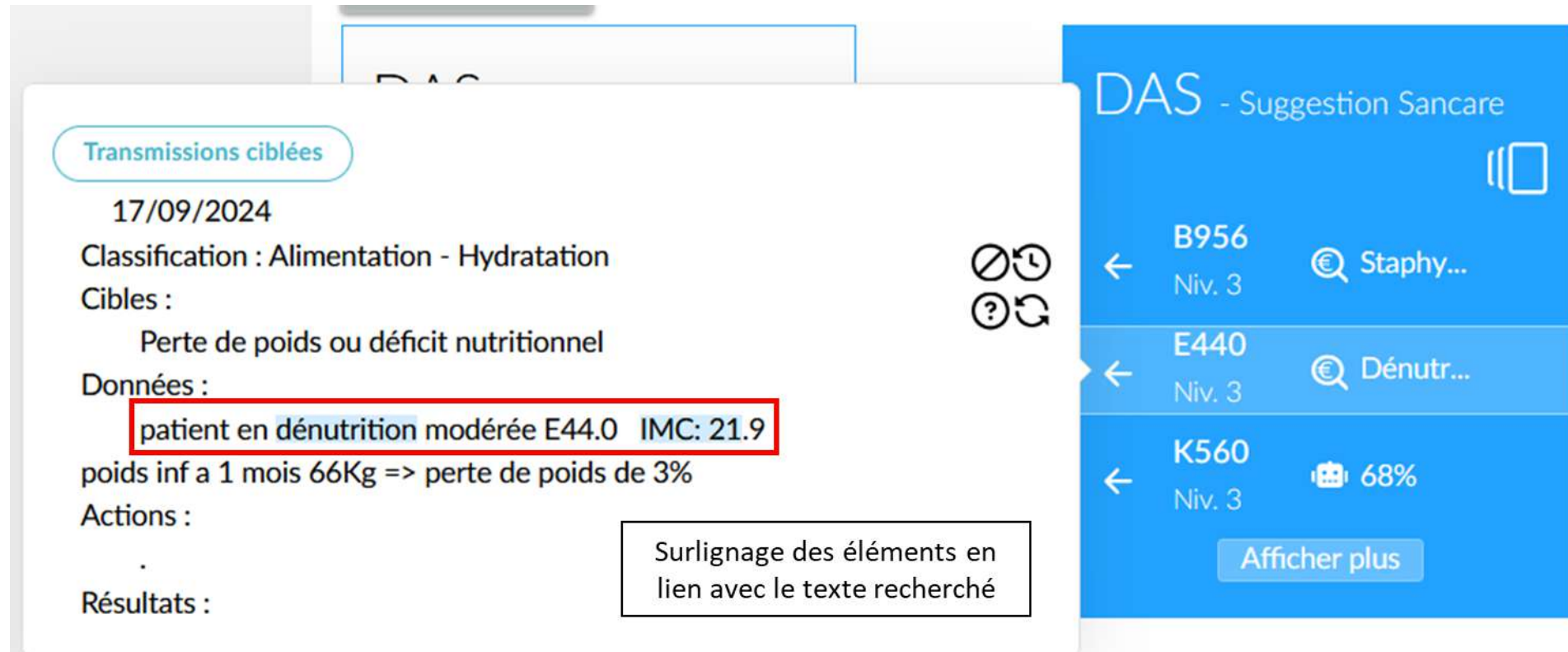




- 1 page par dossier revalorisable



- 1 page par dossier revalorisable
  - Possibilité d'avoir un focus sur la partie du document identifié



The screenshot displays the SANCARE interface. On the left, a patient record for 'DAS' is shown with the following details:

- Transmissions ciblées** (highlighted in a rounded rectangle)
- Date: 17/09/2024
- Classification: Alimentation - Hydratation
- Cibles: Perte de poids ou déficit nutritionnel
- Données: patient en dénutrition modérée E44.0 IMC: 21.9 (highlighted in a red rectangle)
- Additional text: poids inf a 1 mois 66Kg => perte de poids de 3%
- Actions: .
- Résultats: .

On the right, a blue panel titled 'DAS - Suggestion Sancare' displays search results:

- Item 1: B956 Niv. 3 with a magnifying glass icon and the text 'Staphy...'
- Item 2: E440 Niv. 3 with a magnifying glass icon and the text 'Dénutr...'
- Item 3: K560 Niv. 3 with a calendar icon and the text '68%'
- A button labeled 'Afficher plus' is at the bottom.

Navigation icons (back, forward, search, refresh) are visible between the two panels. A callout box at the bottom right states: 'Surlignage des éléments en lien avec le texte recherché'.



# PREDIMED

- Entrepôt de données de santé sur le CHU Grenoble Alpes
  - Création en 2017, partenariat CHUGA / Université Grenoble Alpes / CNRS

- Modèle de « lac de données » :

- Données volumineuses (documents DPI, textes, images, données microbiologiques « \_omiques »), données du PMSI et des données externes (AM (demande en cours), démographiques, environnementales)



A Moreau-Gaudry – Predimed

- Finalités :

- Recherche (accès des chercheurs aux données nominatives)
  - Requêtes textuelles en appui du PMSI effectuées par le DIM à l'aide de l'outil QuickView
    - Moteur de recherche elasticsearch en texte intégral
- Pilotage d'établissement (développement en cours)
- Amélioration de la prise en charge des patients du CHUGA (indicateurs de santé)

Quick View : Recherche dans les documents de type texte

Chargement liste IEP ou IPP

Charger config | Choix index ES : txt\_metadone | Filtrage spécial : Aucun | Charger filtre

Paramètres de la requête (Tous )

- Compte-rendu
- Rés. biologie
- Doc hospit.
- CR anapath
- CR consultation
- Ordo. médic.
- Dossier social
- Antécédents
- CR urgences
- CR opér.
- Transm./Consignes
- Mot synthèse
- Ordonnance
- CR accouch.
- Mot suivi
- Expl. fonction
- CR imagerie
- ECG
- Obs./Exam dln.
- Autres
- CR hospit.
- CR exam.
- CR RCP
- Autres doc.

Mots ou Expressions recherchés : symphysite pubienne | Type expr. : Expression | Opé. App. : OR | Fuzzy 1

Mots ou Expressions recherchés : | Expression  | Opérateur à appliquer : OR | Fuzzy 0

Paramètres de la recherche

Date début : 01 / 01 / 2018 | Sélection des dates

Date fin : 31 / 12 / 2023

Send query | Documents loaded : 151 on eq 151 | Next data | Displayed 1 to 151 | Export selected data | Save context | View last request

Load query

Résultat de la requête

IPP	Venue	Type doc	Format	Date doc	Origine	Lien doc.
		CR de consultation	pdf	2022-09-16	RHUMATO	
		Compte-rendu Bioconcert.	pdf	2022-07-08	RHUMATO	
		CR de consultation	pdf	2018-03-12	RHUMATO	
		Compte-rendu Imagerie	pdf	2023-09-27	SECTEUR ONCOLOGIE	
		EMS	pdf	2019-10-03	HAD	
		HAD : Fiche d'hospitalisation J0 CHUGA	pdf	2019-09-27	HAD	
		Staff	pdf	2019-08-23	MED INFECT	
		Méd. Infectieuse : Fiche d'hospitalisation J0 CHUGA	pdf	2019-08-09	MED INFECT	

Fragment de texte contenant les expressions recherchées

réveillant la nuit. Pubalgie chronique en rapport avec une **symphysite pubienne** bien visualisée à l'IRM Pas de signe extra-rhumatologique (cutané)

Extrait du courrier avec résultat positif

Load query : Intégration de requêtes plus complexes utilisant le langage domain-specific language (DSL)  
Fichier d'import au format .json

Ouverture du document original

- Utilisation PMSI :
  - RAAC, vérification du codage de certaines prises en charge...
  - Demandes des services (recherche ou activité) lorsque CIM10 non pertinentes
  
- Limites rencontrées de l'outil Quickview
  - Les documents ne sont pas forcément rattachés à un numéro de séjour
    - Problématique Easily
  - Utilisation de la date de création du document qui peut être différente des dates de séjour
  - Automatisation impossible : extraction de PREDIMED puis croisement avec les données PMSI
    - Limite d'extraction : 500 lignes

# Outil interne CHUGA

- Utilisation du contenu des bases du DPI (Easily) pour de la recherche textuelle à visée d'optimisation
- Objectifs de l'outil :
  - Facilité d'utilisation
    - Identification rapide des faux positifs et des optimisations potentielles
  - Facilité d'exécution
  - Facilité de paramétrage des CQ, de mise à jour et de maintien de l'outil



- Logiciels utilisés

- Pré requis : Accès aux bases de données Easily (lecture seule, copie de la base)
- Script R : requête SQL sur les bases Easily puis traitement des données et automatisation
- Export d'un fichier Excel de résultat pour traitement

- Processus





- Nécessité d'un retraitement des données exportées d'Easily avant d'effectuer la recherche textuelle
- 3 niveaux de complexité :
  - 😊 Données textuelles directement en base
    - Transmissions ciblées, fiches diet, mot de suivi, pancarte...
  - 😐 Données XML au sein des tables SQL
    - Fiches d'observations médicales, macrocibles, fiches Easily CRO...
  - 😞 Données binaires
    - Courriers (source Word) et PDF

# Les données des bases Easily : données directement en base

⏏
⚠ CONFUSION 1

DONNÉES	ACTIONS	RÉSULTATS
M G est confus. Il cherche un anneau qui lui serait précieux	Réassurance sur l'inventaire réalisé à l'entrée et les objets précieux conservés en coffre	M G ne semble pas totalement convaincu mais se calme

📅 22/09/2024 04:02

ci_cat_libelle	fi_id	fi_creation_date	fi_donnees	fi_actions	fi_resultats
Confusion	XXXXXX	2024-01-01 00:12:34.5	M G est confus. Il cherche un anneau qui lui serait précieux	Réassurance sur l'inventaire réalisé à l'entrée et les objets précieux conservés en coffre	M G ne semble pas totalement convaincu mais se calme

**Compte-rendu de soins diététiques**

Date	Diététicien
<b>Diagnostic diététique</b>	
Poids mesuré à 30kg pour 1m40 soit un IMC à 15	
Monsieur G se nourrit exclusivement de poisson lorsqu'il arrive à les attraper. Fort déséquilibre alimentaire. Refuse les aliments cuits.	
> Dénutrition sévère sur état de maigreur avec déséquilibre alimentaire	
<b>Stratégie de prise en charge</b>	
Standard, 1 CNO lacté	

fi_id	date_mot_prise_e n_charge_cr	intervenant cr_txt	mot_prise_en_charge_txt_cr	consignes_prise_e n_charge_inclu
XXXXXX	11/03/2024	XXXXXX	Poids mesuré à 30kg pour 1m40 soit un IMC à 15 Monsieur G se nourrit exclusivement de poisson lorsqu'il arrive à les attraper. Fort déséquilibre alimentaire. Refuse les aliments cuits. > Dénutrition sévère sur état de maigreur avec déséquilibre alimentaire	Standard, 1 CNO lacté

ADMISSION

Admission

Entree programmee

Date heure arrivee unite

Unite fonctionnelle MED POLYVALENTE

Etablissement CH DU G

Antecedents significatifs Troubles de la personnalité

Informations observations cliniques Patient hospitalisé en état d'incurie avancé, forte déshydratation, dentition altéré. Bilan AEG

1- Récupération du contenu XML

```
<motif_hospitalisation>Bilan AEG</motif_hospitalisation>
<entree_programmee>false</entree_programmee>
<entree_urgence>false</entree_urgence>
<date_heure_arrivee_unite>00</date_heure_arrivee_unite>
<heure_arrivee is_null="true"/>
<unite_fonctionnelle>MED POLYVALENTE</unite_fonctionnelle>
<etablissement>CH DU G</etablissement>
<circonstances_arrivee_mode_transport>Arrivé par le SAMU</circonstances_arrivee_mode_transport>
<precaution_complementaire is null="true"/>
<antecedents_significatifs>Troubles de la personnalité</antecedents_significatifs>
<informations_observations_cliniques>Patient hospitalisé en état d'incurie avancé, forte déshydratation, dentition altéré. Bilan AEG</informations_observations_cliniques>
<specialite_fiche_intro>
```

2- Extraction des informations utiles du fichier XML

fi_id	antecents_significatifs	informations_observations_cliniques
XXXXXX	Troubles de la personnalité	Patient hospitalisé en état d'incurie avancé, forte déshydratation, dentition altéré. Bilan AEG

- Variables de type BLOB (Binary Large Object) dans les bases SQL
  - Courriers, documents word...
- Génération en local du fichier (docx, pdf...) à partir des données binaires
- Chargement du fichier dans R
  - Suivant l'utilisation, peut nécessiter un retraitement pour les champs non pertinents (en têtes, identité du patient, information sur le service...)
- Réalisation de la requête de recherche textuelle



Catégorie	Terme recherché	Durée de séjour minimale	Sévérité maximale	Codes CIM10 attendus
Fécalome	f.calome	4	2	K564
Confusion	confu	3	1	F05   R410
Confusion	d.sorientation	3	1	F05   R410
Dénutrition	d.nutr	4	2	E40   E41   E42   E43   E440

- **Catégorie** : plusieurs recherches possibles dans une seule catégorie
- **Terme recherché** : utilisation des expressions régulières pour essayer de prendre en compte les différentes orthographes
  - Exemple : le « é » de fécalome est remplacé par un « . » qui permet d'accepter n'importe quel caractère
- **Durée de séjour minimale et sévérité maximale**
  - Permet de filtrer les séjours qui ne bénéficieraient pas de l'ajout du code
- **Codes CIM10 attendus** :
  - Donne une indication sur les codes attendus et permet d'exclure les séjours qui ont déjà ces codes



- Gestion des fautes de frappe
  - Actuellement : manuellement en supprimant les caractères à risques, en se limitant à la racine du terme...
  - Existence d'algorithmes de recherche approximative (fuzzy search) qui pourraient être implémentés
    - Prennent en compte l'ajout / la suppression / la substitution d'un nombre de caractère
- Exclusion spécifiques sur certains points : « chute d'escarres » qui ressortent avec les escarres, « globes oculaires » qui ressortent dans la recherche du terme « globe » dans les RAU...
- Négation :
  - Actuellement : exclusion manuelle
    - Termes spécifiques associés au thésaurus : « pas de », « absence de signes de »...
  - Difficultés sur les négations plus complexes : « absence de [...], de détresse respiratoire... »
  - A terme : algorithme plus poussé, machine learning pour repérer les négations ?



- Les séjours positifs sur une recherche textuelle sont croisés avec les données PMSI
  - Filtre en fonctions des informations renseignées dans le thésaurus
  - Durée de séjour
  - Niveau de sévérité
  - Exclusion si le diagnostic recherché (ou un code de la même catégorie) déjà codé



- **Objectif** : simplicité du traitement des résultats
  - Faux positifs identifiés facilement
  - 1 fichier Excel / pôle
  - La colonne "Commentaire" est réutilisée sur les mois d'après pour identifier les résultats déjà traités par les MIM
- Périodicité : mensuelle

IEP	UM	DUREE	GHM	VALO	INFO	SOURCE	DATE	TEXTE	COM DIM
XXXXXXXXXX	XXXX		8 01C091	8 063€	Confusion - F05 R410	Transmissions ciblées	XXXXXXXXXX	<p><b>TC Données</b> M G est <b>confu</b>s. Il cherche un anneau qui lui serait précieux</p> <p><b>TC Actions</b> Réassurance sur l'inventaire réalisé à l'entrée et les objets précieux conservés en coffre</p> <p><b>TC Résultats</b> M G ne semble pas totalement convaincu mais se calme</p>	
XXXXXXXXXX	XXXX		8 01C091	8 063€	Dénutrition sévère - E40 E41 E42 E43 E440	Fiche Diet	XXXXXXXXXX	<p><b>Mot Diet</b> Poids mesuré à 30kg pour 1m40 soit un IMC à 15</p> <p>Monsieur G se nourrit exclusivement de poisson lorsqu'il arrive à les attraper. Fort déséquilibre alimentaire. Refuse les aliments cuits.</p> <p>&gt; <b>Dénutrition sévère</b> sur état de maigreur avec déséquilibre alimentaire</p> <p><b>Consignes Diet</b> Standard, 1 CNO lacté</p>	
XXXXXXXXXX	XXXX		8 01C091	8 063€	Hypovolémie - E86	Macro cible	XXXXXXXXXX	<p>Patient hospitalisé en état d'incurie avancé, forte <b>déshydr</b>atation, dentition altéré. Bilan AEG</p>	23

- Techniquement abordable
  - Sous réserve d'accès au DPI
  - Réalisation sous R mais transposable facilement sur d'autres langages
- Exécution manuelle mensuelle donc pas de résultats au fil de l'eau
  - Temps d'exécution optimisable : 10-15 minutes en fin d'année
- Optimisation du script
  - Actuellement : Extraction Easily → recherche textuelle → croisement CORA pour ne sélectionner que les dossier éligibles
  - Inverser le processus pour n'extraire que les dossiers utiles d'Easily à partir d'une liste de dossiers PMSI éligibles ?
- Problématique Easily :
  - IEP non utilisables dans beaucoup de cas
  - Nécessité de faire une jointure avec les données PMSI sur la date et l'IPP → lourd

- Connaissance des bases d'Easily et de leur fonctionnement
- Utilisation pour des requêtes ponctuelles
  - A visée d'optimisation : SMR (fiches assistantes sociales), RPM en obstétrique
  - Activité de praticiens : ex : activité assistant comme aide-opérateur en chirurgie (variable présente dans le CRO uniquement), activité d'un kiné sur un service...

# Conclusion



- Utilité de la recherche textuelle dans l'optimisation PMSI
  - Contrôles ciblées orientant la recherche des CMA : moins à l'aveugle que dans des contrôles de sous valorisation
  - Mais ne peut pas être exhaustif (4800 codes CMA)
- Attention à l'inflation du nombre de CQ à réaliser intégrés dans le PAQ
  - Produire ce qu'on est capable d'absorber (MIM, TIM)
- Complète mais ne remplace pas les contrôles de base PMSI
  - Sous valorisation, contrôle des DP, qualité de codes CIM10, biologie...
- Et demain : on appelle ChatGPT ?

Merci pour votre attention