

KaliDOS

Qualité des données du Sudoc

Réalisé par :

ADBI Karim

AGAGNA Radjaa

DELFANTI Maxence

FIUME Dominique

LEANO MARTINET Michelle

MONTEIX Yann

I. Enjeux & Objectifs



Contexte



Qualité du catalogage à la BU

Chaque nouvelle œuvre (livre, DVD, thèse, etc.) est cataloguée par la BU selon des normes nationales :

Besoin d'un **contrôle qualité**

- Utilisation de services distants
- Vérification automatique de règles
- Détection des mentions d'entités non référencées



Contexte

```
<datafield tag="328" ind1="" ind2="">
  <subfield code="d">
    2021
  </subfield>
  <subfield code="a">
    SuDoC
  </subfield>
</datafield>
```



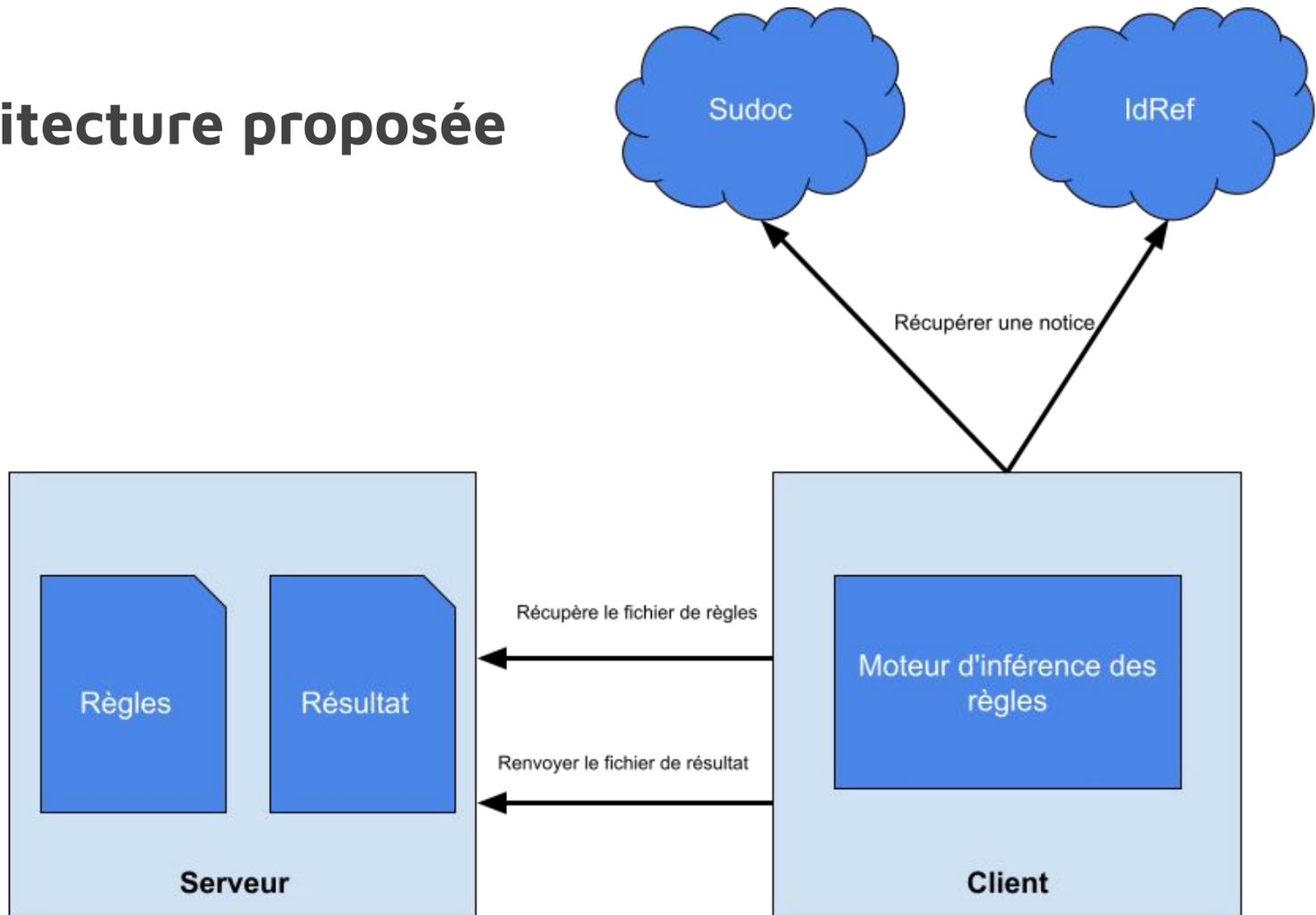
```
{
  "regex": "^[0-9]{4}$",
  "code": "d",
  "message": "Zone 328$d doit contenir
uniquement l'année de soutenance",
  "number": 328,
  "index": 3,
  "numRuleExcell":122
}
```



```
{
  "regex": "^(?!SuDoC|SUDOC).*$",
  "code": "d",
  "message": "Zone 328$a ne doit pas
être égal à SuDoC",
  "number": 328,
  "index": 3,
  "numRuleExcell":122
}
```



Architecture proposée





SUDOC

+ Efficace

- Jeu de règles
obsolète et non
modulable

- Ergonomie

- Export des résultats
impossible

← → ↺ 🏠 | domybiblio.net/check_sudoc/check_sudoc.php?action=do_process&id_s=0 | 67 %

CheckSUDOC Accueil DoMyBiblio UnivDoc SUDOCAnalyse SUDOCLocate

CheckSUDOC

Contrôle qualité de vos notices du SUDOC

Nous ne comptons que les zones 60X provenant de l'indexation Rameau mais affichons l'ensemble des zones 60X dans le détail.
Vous pouvez envoyer vos commentaires et suggestions à [contact\[@\]domybiblio.net](mailto:contact[@]domybiblio.net).

2 identifiant(s) à traiter.

[Retour](#)



[PPN 249033348]

● **Manuel de soins palliatifs**

PPN : 249033348

Années : 100 = 2019 210_d=

Champs 181/182 :

Champ 200 : 200_a Aucun double espace dans le titre. - 200_b 0

Champ 210 : 210_a : - 210_c : - 210_d :

Champ 214 : 214_a : Pas de Mention latine S.I. - 214_c : Pas de Mention latine S.n. - 214_d : Pas de Mention latine S.d.

Champ 225 : 0

Champ 305 : 0

Champ 410 : 0

Champ 461 : 0

Champ 606 : Aucune anomalie constatée

Champ 700 : Aucune anomalie relative au \$3 Aucune anomalie relative au \$4 Le contenu du code fonction en \$4 ne correspond pas à ce qui est attendu en 700

Champ 701 : Il manque un \$3 en 701 Aucune anomalie relative au \$4 Le contenu du code fonction en \$4 ne correspond pas à ce qui est attendu en 701



[PPN 249032554]

● **Soins palliatifs [en 54 notions]**

PPN : 249032554

Années : 100 = 2020 210_d=

Champs 181/182 :

Champ 200 : 200_a Aucun double espace dans le titre. - 200_b 0

Champ 210 : 210_a : - 210_c : - 210_d :

Champ 214 : 214_a : Pas de Mention latine S.I. - 214_c : Pas de Mention latine S.n. - 214_d : Pas de Mention latine S.d.

Champ 225 : 1

Champ 305 : 0

Champ 410 : 1

Champ 461 : 0

Champ 606 : Aucune anomalie constatée

Champ 700 : Aucune anomalie relative au \$3 Aucune anomalie relative au \$4 Le contenu du code fonction en \$4 ne correspond pas à ce qui est attendu en 700

Champ 701 : Aucune anomalie relative au \$3 Aucune anomalie relative au \$4 Le contenu du code fonction en \$4 ne correspond pas à ce qui est attendu en 701

II. Organisation

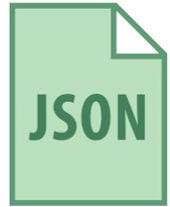


Rôles

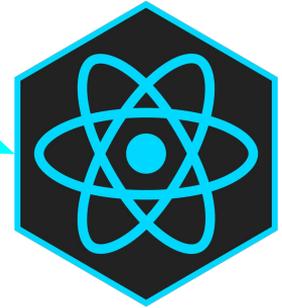
	Poste	Domaine	Tâches
Maxence	chef de projet	front/back	Saisie des règles et interface de vérification, debug, tests unitaires, documentation
Yann	lead dev	back	Saisie des règles, Implémentation des règles, déploiement, debug, tests unitaires, doc
Karim (départ fin nov)	dev	back	Saisie des règles, Implémentation des règles, tests unitaires
Radjaa	dev	front	Interface de règles, rapport de qualité, doc
Michelle	dev	front	Interface de règles, doc
Dominique	dev	front/back	Historique des notices, rapports excel, doc



Technologies

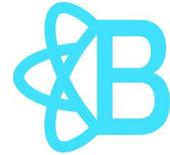
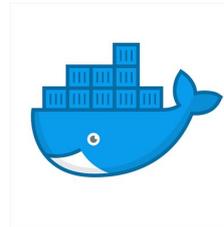


node JS
express



{ } model_regles_tries.json

{ } notices_erreurs.json





Modélisation des règles (1)

10 types de règles différentes

- **Matching :**
 - 008 doit contenir "x3"
- **Structurelle :**
 - 309 ne doit pas être présente
- **Dépendance :**
 - 029\$b, les 4 premières positions doivent être identiques à celles de 328\$d
- **Conditionnelle de Matching :**
 - Si 214 #0 et \$d, doit commencer par "DL espace" ou "[" ou un chiffre
- **Conditionnelle Structurelle :**
 - Si 008 commence par Aa, 215 \$a et \$d obligatoires
- **Conditionnelle de Dépendance :**
 - Si 225 ind1=0 \$a est différent du 410\$t



Modélisation des règles (2)

- **Référentiel :**
 - Si 602\$a, vérifier dans ldref, à partir de l'identifiant \$3, que 008 commence par Ta
- **Précédence :**
 - 6XX \$a et \$2fmesh, \$a doit être précédé d'un \$3
- **Compte :**
 - Si plusieurs 101\$d, il doit y avoir autant de 330
- **Ordonnement :**
 - Si plusieurs zones 214, doivent respecter l'ordre des chiffres de l'ind2



Phasage

	S1	S2	S3	S4	S5	S6
Saisie des règles	■					
Implémentation des règles	■					
Tests unitaires			■			
Interface de vérification	■					
Interface de règles		■				
Rapports Excel		■				■
Historique des notices			■			
Debug			■			
Documentation						■
Déploiement	■					

III. Bilan



Saisie

Choix du jeu de règles

Générale

Générale

Monographie électronique

Thèses et Mémoires (soutenance)

Thèses mémoires (reproduction)

Monographies imprimées et autres documents

Saisie des identifiants

Saisie des identifiants

Choix du jeu de règles

Générale

Identifiants (1 PPN par ligne) :

Saisissez votre liste de PPN, ou faites glisser un fichier dans la zone

Envoyer

Vérification

Nombre de PPN testés : 5

Nombre total d'erreurs : 35

Nombre de PPN inexistants : 0

Interface de vérification

📄 Descriptif complet

📄 Double entrée

Erreurs par PPN

Recherche...



169450554

9

169450570

5

169450589

9

169450597

7

169451119

5

< 1 >

Détail des erreurs par PPN

Message d'erreur	Zone	Sous zone	ID Excel
Zone 105 à compléter	105	a	20
Zones 6XX : \$2 mal orthographié	606	2	83
100\$a Pos. 10-13 = 214\$d	100 , 214	a , d	2
Mention d'index incohérente : vérifier zones 105 et 320	320	a	15
Si 105 \$a 3e position en commençant par la fin =0 alors il ne doit pas y avoir une 320 avec la mention "Index"	320	a	17



Historique

Historique des notices erronées

Supprimer l'historique 

Notices erronées

23/02/2021 09:23:46	6
23/02/2021 09:25:32	6
23/02/2021 09:26:28	5
23/02/2021 10:29:14	2

Détail des notices

Relancer le contrôle

Supprimer la notice 

169450546

169450554

Règles

Ajouter une règle

x

Interface

Chisissez un type de règle :

- Matching
- Dépendances
- Conditionnels Matching
- Conditionnels Dependance
- Conditionnels Structurel
- Structurels
- Références
- Ordonnancement
- Compte
- Precedence

Jeu de règle

Rechercher

Type de document

- Monographies et autres
- Monographies et autres
- Monographies et autres

Ajouter une règle

Règle de type : matching

Type de document*
Générale

Identifiant Excel*
1

Quantité de datafields*
Un

Zone*
230

Sous Zone
a

Création du regex*
Regle a utiliser*

Action

Suivant

Matching

Ce type de règle permet d'imposer une contrainte sur la chaîne de caractère d'une Sous Zone si elle existe. Nous rappelons l'exemple : Le champ (Zone\$Sous Zone) 230\$a ne doit pas contenir 'Mo'. Il est également possible d'inclure plusieurs champs dans une même règle et plusieurs motifs à matcher. Par exemple : 700\$4 et 701\$4 ne doivent pas contenir les valeurs 020, 050, 060.

Ajouter une règle

x

Règle de type : matching

Type de document*

Générale

Identifiant Excel*

1

Quantité de datafields*

Un

Zone*

230

Sous Zone

a

Création du regex*

Regle a utiliser*

Action



if

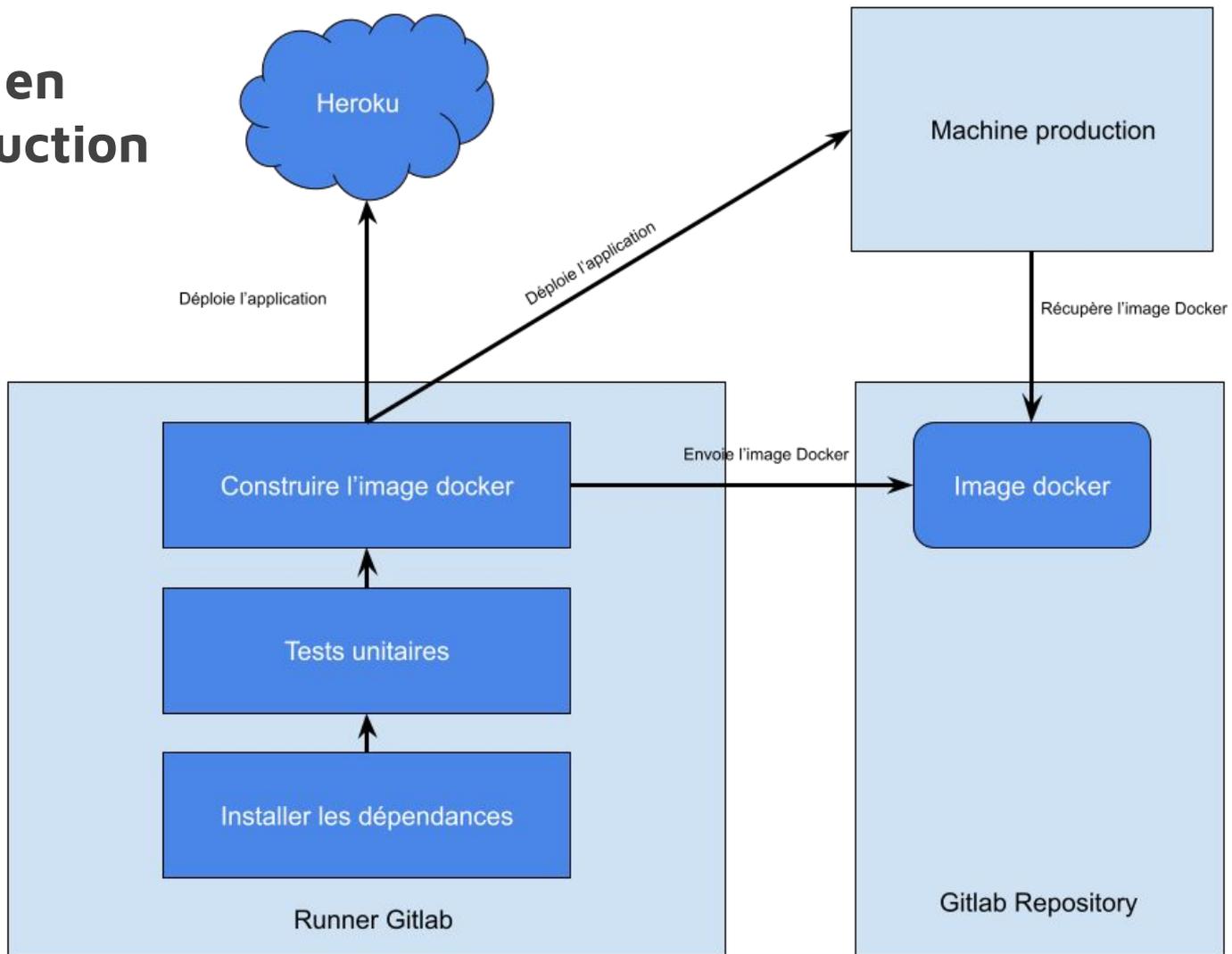


ice



Suivant

Mise en production





Retour d'expérience

Interne à l'équipe

- Gestion de projet
- Communication
- Investissement personnel et motivation



Perspectives & Évolution

- Bon retours utilisateurs
- Perspective d'une application à l'échelle nationale (ABES)

MERCI