Présentation de la plateforme DeepCorp adaptée à ESLO

BRAS Enzo - année 2023
ASTN/LLL/CNRS

Sommaire

- Les besoins
- Analyse de l'existant
- Nouveau site
  - CouchBase
  - Spring Boot
  - Wordpress
- Amélioration / Projection
- Bilan
  - Problèmes rencontrés
  - Ressenti

Les besoins

Le services ASTN a entrepris le développement de plusieurs plateformes d'outillage. Une des premières plateformes développées est AlimCorp. Alimcorp couvre les phases de création, de collecte et d'enrichissement de Corpus.

- ➔ La solution actuelle web du site d'ESLO qui permet aux chercheurs de naviguer dans le Corpus est devenue obsolète et ne permet pas d'évolution majeure.
- ➔ La solution retenue par le service ASTN est donc le développement d'une plateforme dédiée à l'extraction et aux traitemets sur les données de ces corpus, pleforme nommée DeepCorp.
- ➔ Nouvelle plateforme qui doit permettre d'offrir des fonctionnalités de traitements et d'analyses de haut niveau aux chercheurs à partir des données collectées par AlimCorp.

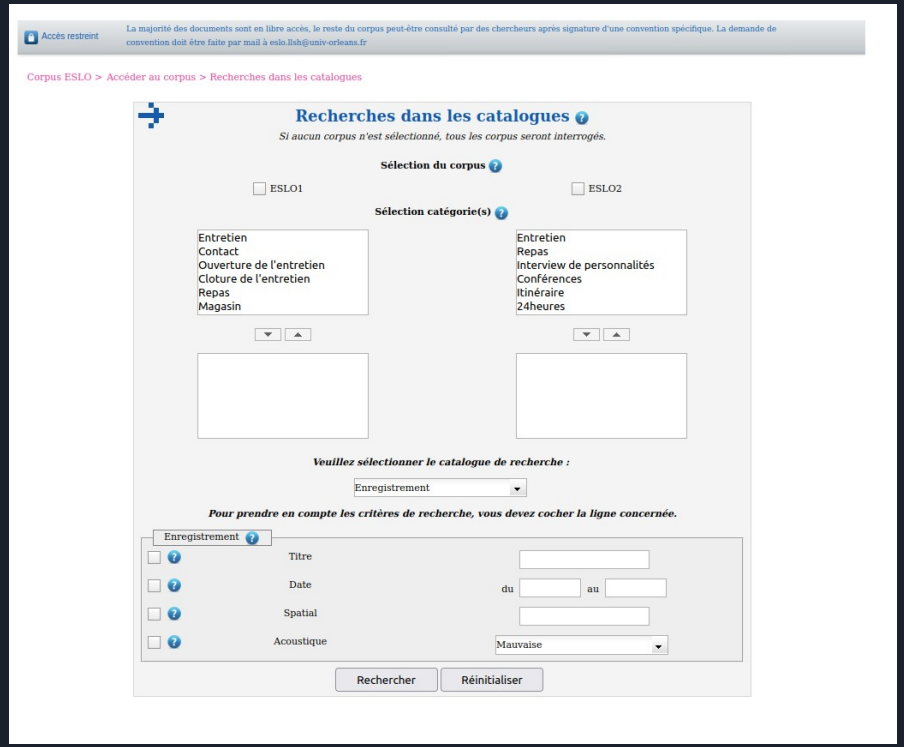
Site associé au Corpus Eslo

perte de la maitrise

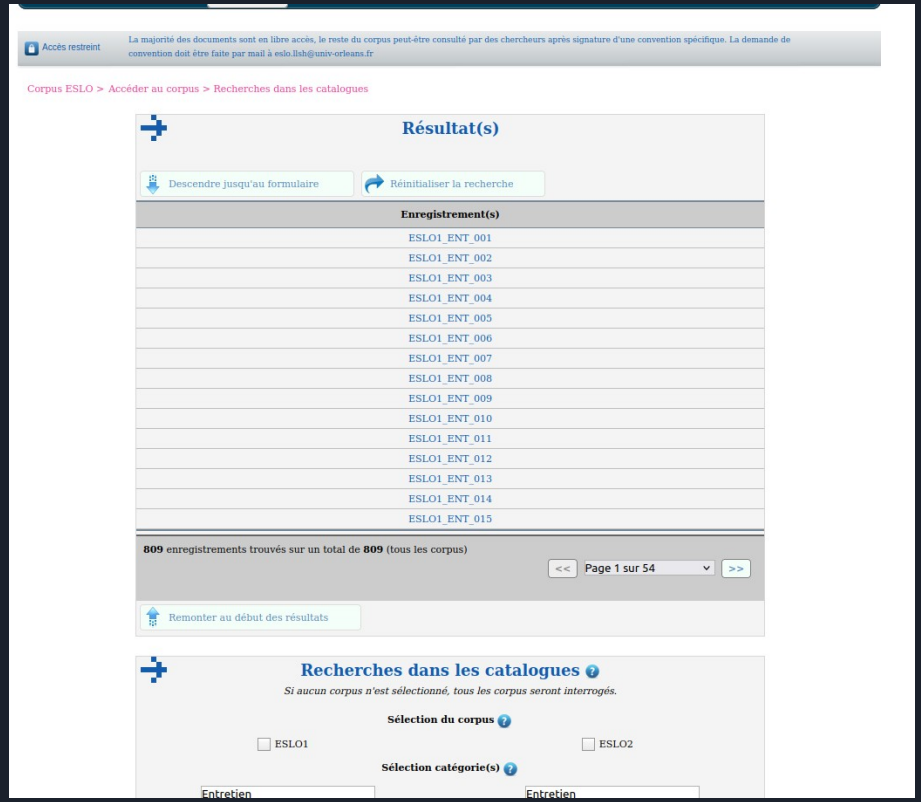
joomla

Difficulté à faire évoluer

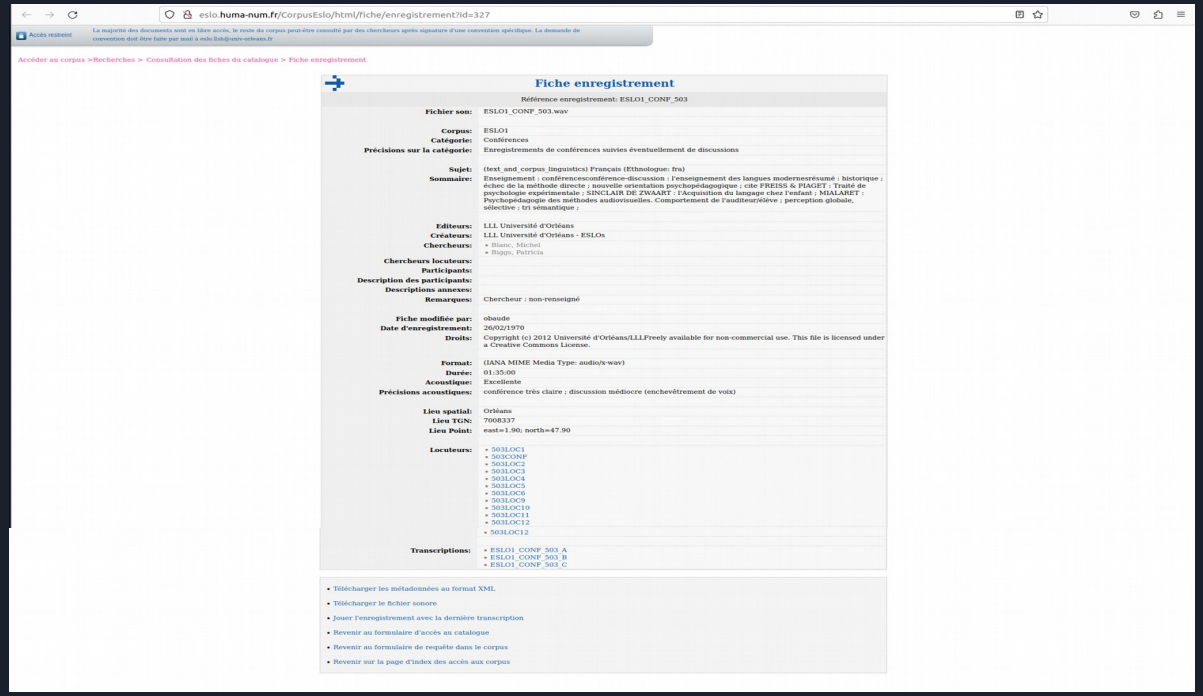
Page formulaire de saisie de critère de recherche



Affichage résulat d'une recherche



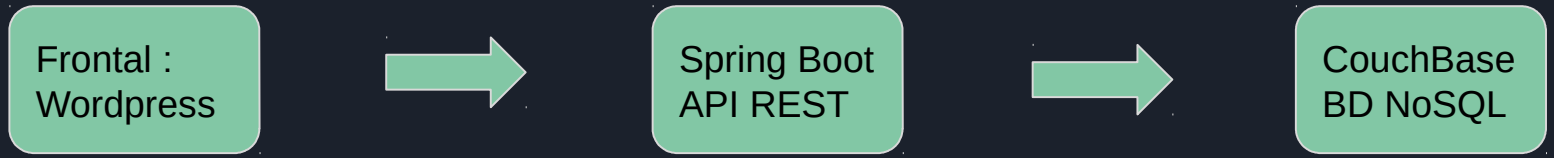
Accés aux détails d'un enregistrement



Nouveau site DeepCorp

Projet : prototyper une solution d'extraction et d'analyse à partir du modèle Eslo.

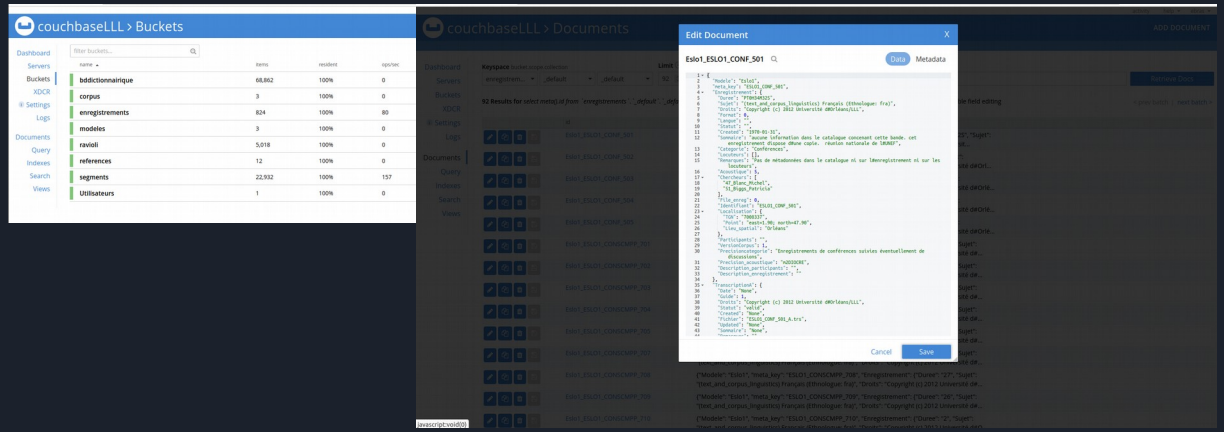
La solution s'appuie sur une architecture multi-tiers constituée :
- Frontal sous wordpress
- Web service (spring boot)
- Backend : couchBase et Kubernetes



Choix du CMS était imposé par le service ASTN.

Backend / CouchBase

Le cluster CouchBase est une solution de base de données Nosql de type document. Les objets contenus dans cette base sont organisés en Buckets.Les documents ainsi stockés sont de nature différentes et au format JSON.



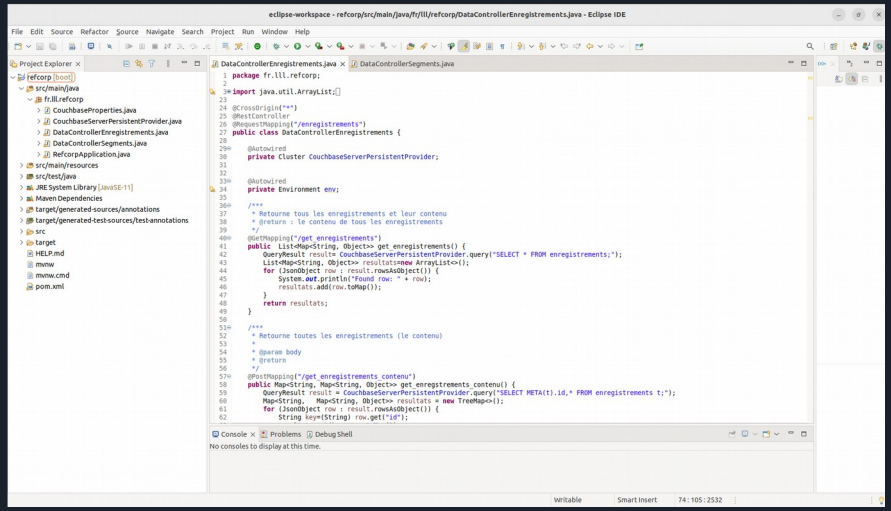
Flexibilité

Performance

Scalabilité

Backend / Spring Boot

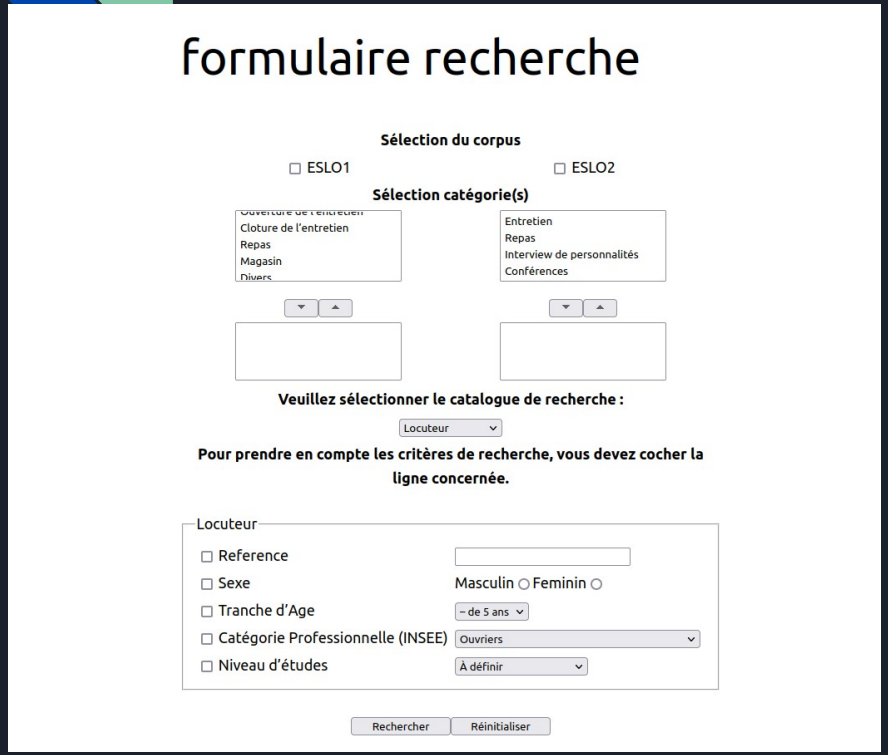
- Ce web service a pour objectif de desservir et d'alimenter le frontal en effectuant la demande d'exécution des requêtes au cluster CouchBase. Il a pour objectif de mutualiser les demandes car ASTN a plusieurs projets qui ont besoin de ce service. De plus, le fait de séparer la logique de stockage du frontal permet le cas échéant de changer de solution si besoin.

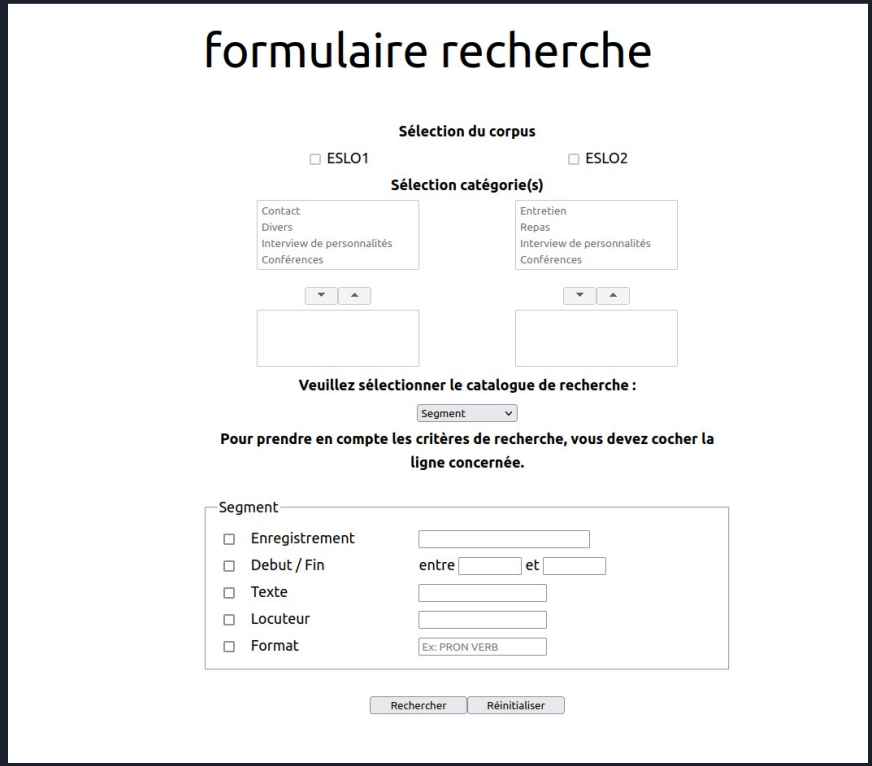


Travail à effectuer

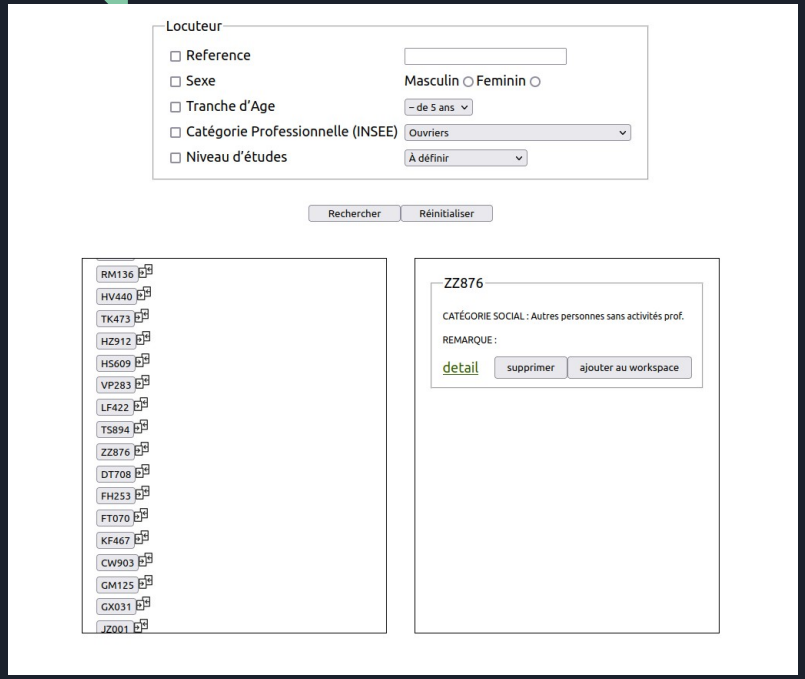
- Mon objectif est de développer principalement la partie Extraction, sachant que les autres pages ne sont que informations statiques que l'on complétera par la suite.
- Pour la partie extraction, je vais reproduire l'existant et eprmettre des extractions sur les types de données enregistrements, transcriptions et locuteurs.
- Je vais ensuite rajouter les possibilités de traitements aux autres types de données de type segments et d'offrir la possibilité d'explorer également les descriptions plus précises des segments.

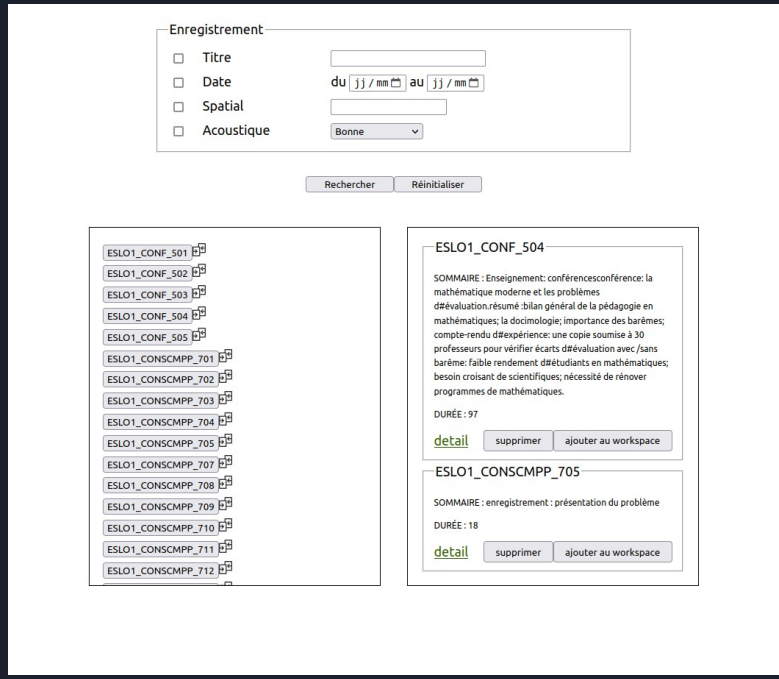
Frontal / page formulaire



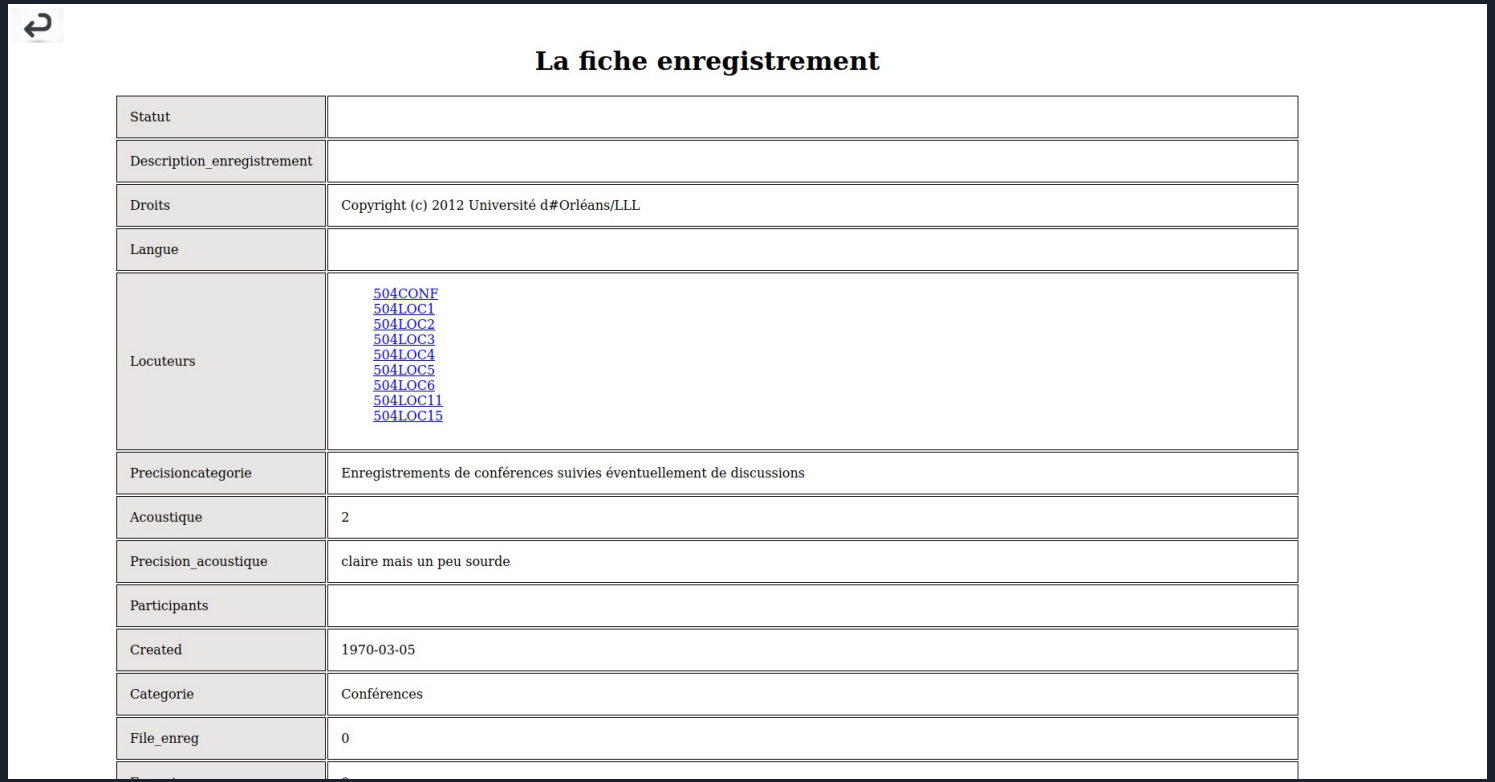


Nouveau site / page formulaire





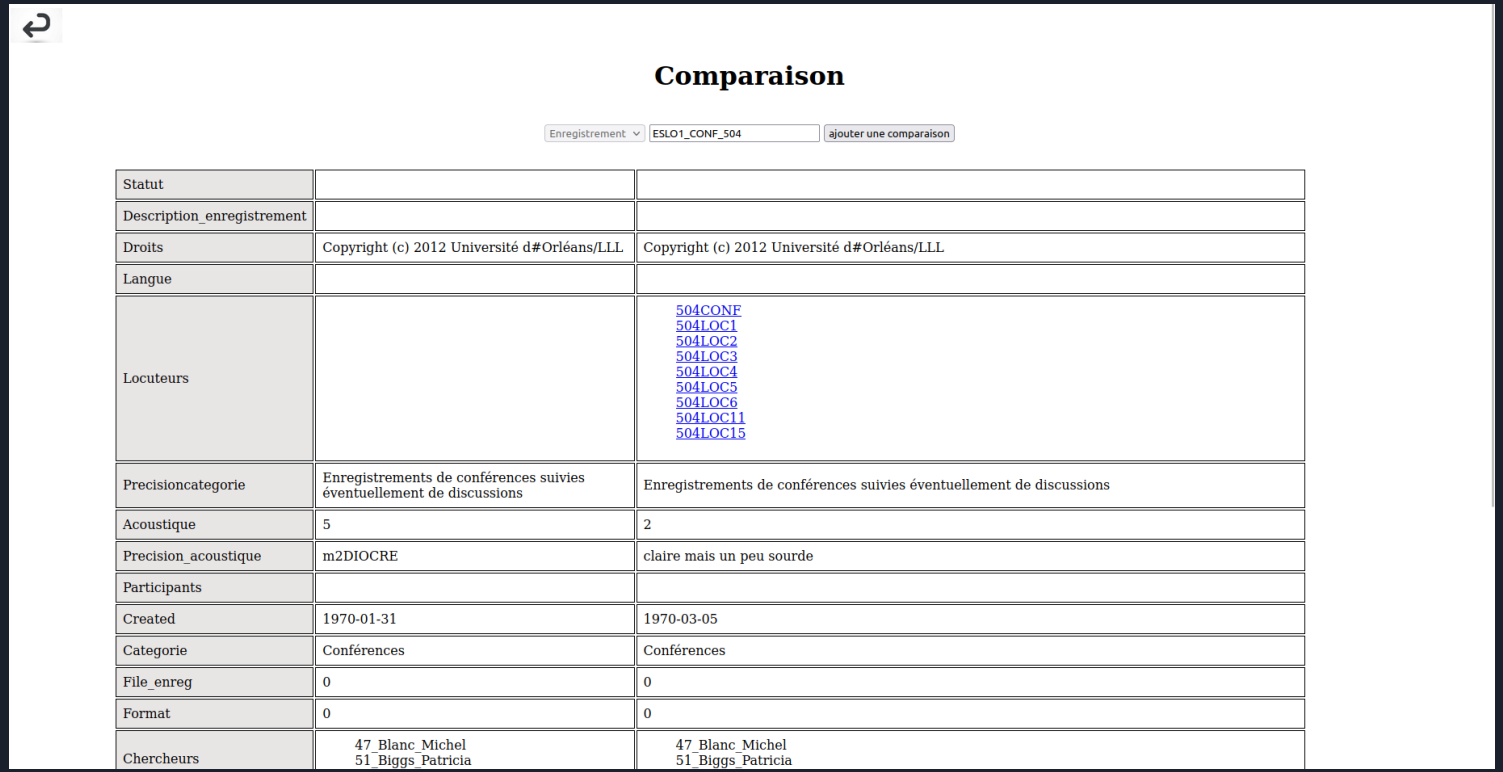
Nouveau site / page détail



La fiche enregistrement

Statut	
Description_enregistrement	
Droits	Copyright (c) 2012 Université d#Orléans/LLL
Langue	
Locuteurs	504CONF 504LOC1 504LOC2 504LOC3 504LOC4 504LOC5 504LOC6 504LOC11 504LOC15
Precisioncategorie	Enregistrements de conférences suivies éventuellement de discussions
Acoustique	2
Precision_acoustique	claire mais un peu sourde
Participants	
Created	1970-03-05
Categorie	Conférences
File_enreg	0

Nouveau site / page comparaison



Comparaison

Enregistrement | ESLO1_CONF_504 | ajouter une comparaison

Statut		
Description_enregistrement		
Droits	Copyright (c) 2012 Université d#Orléans/LLL	Copyright (c) 2012 Université d#Orléans/LLL
Langue		
Locuteurs		504CONF 504LOC1 504LOC2 504LOC3 504LOC4 504LOC5 504LOC6 504LOC11 504LOC15
Precisioncategorie	Enregistrements de conférences suivies éventuellement de discussions	Enregistrements de conférences suivies éventuellement de discussions
Acoustique	5	2
Precision_acoustique	m2DIOCRE	claire mais un peu sourde
Participants		
Created	1970-01-31	1970-03-05
Categorie	Conférences	Conférences
File_enreg	0	0
Format	0	0
Chercheurs	47_Blanc_Michel 51_Biggs_Patricia	47_Blanc_Michel 51_Biggs_Patricia

Suivi et tuning / logs

```
mysql> select * from wp_logs_exec;
+---------+---------------+------------------------------------------------------------------------------------------------------------------------------------------
------------------+--------------------+
| user_ID | catalogue     | requete
                  | date               |
+---------+---------------+------------------------------------------------------------------------------------------------------------------------------------------
------------------+--------------------+
|       1 | enregistrement |  ( Modele = 'Eslo1' )  AND  ( Enregistrement.Categorie = 'Entretien' OR Enregistrement.Categorie = 'Magasin' )  AND Enregistrement.Acoustique = 4
                  | 2023-06-01 09:43:46 |
|       1 | locuteur       |  ( enregistrements.Modele = 'Eslo1' OR enregistrements.Modele = 'Eslo2' )  AND locuteur.Sexe = 1 AND locuteur.Tranche like '40/45' AND locuteur.categ_socio like 'Ouvriers'
                  | 2023-06-01 09:44:15 |
|       1 | transcription  |  ( enregistrements.Modele = 'Eslo1' OR enregistrements.Modele = 'Eslo2' )
                  | 2023-06-01 09:44:32 |
|       1 | segment        | 0=0
                  | 2023-06-01 09:44:51 |
|       1 | enregistrement |  ( Modele = 'Eslo1' )  AND  ( Enregistrement.Categorie = 'R◆union' OR Enregistrement.Categorie = 'Repas' OR Enregistrement.Categorie = 'Appel t◆l◆phonique' )  AND Enregistrem
ent.Acoustique = 2 | 2023-06-01 09:45:33 |
|       1 | enregistrement |  ( Modele = 'Eslo1' OR Modele = 'Eslo2' )  AND Enregistrement.Identifiant like '%ESLO1_CONF_505%'
                  | 2023-06-01 09:46:01 |
+---------+---------------+------------------------------------------------------------------------------------------------------------------------------------------
------------------+--------------------+
6 rows in set (0.00 sec)
```

Afin d'offrir un service de qualité, un système de mise en place de logs permet de notifier les requêtes effectuées et leur temps d'exécution.
Ce relevé de métriques va nous permettre de suivre la capacité du backend à répondre aux différentes sollicitations.

Amélioration / Projection

- ➔ Ergonomie
- ➔ Gestion des erreurs
- ➔ Ajout de graphique dans la page comparaison
- ➔ Faire un tri dans les éléments affichés dans la page détail

Bilan / Problème rencontré

- ➔ développement de l’API REST
- ➔ compréhension de WordPress

Bilan / Ressenti

- ➔ développement de nouvelles compétences
- ➔ des choses à améliorer

Merci pour votre écoute