

# Documentation de l'IHM DJ-DDCOL

## Sommaire

<a href="#">Documentation de l'IHM DJ-DDCOL</a>	1
<a href="#">Objectifs</a>	2
<a href="#">Alignement du référentiel des personnes physiques de la Direction Juridique (DJ) avec celui de la Direction des Collections (DDCOL)</a>	2
<a href="#">Objectif général</a>	2
<a href="#">Structure des jeux de données initiaux</a>	2
<a href="#">Vérification des alignements réalisés avec Talend</a>	3
<a href="#">Création de nouveaux alignements</a>	4
<a href="#">Interface réalisée</a>	4
<a href="#">Choix</a>	4
<a href="#">Langage de programmation</a>	4
<a href="#">Outils</a>	4
<a href="#">Structure des données</a>	5
<a href="#">Structure de l'interface</a>	6
<a href="#">Description générale de l'interface</a>	6
<a href="#">Description précise des URLs</a>	7
<a href="#">Fonctionnalités</a>	9
<a href="#">Validation d'un alignement</a>	9
<a href="#">Alignement à réaliser</a>	10
<a href="#">Recherche dans les jeux de la DJ et de la DDCOL</a>	11
<a href="#">Annexes</a>	12
<a href="#">Annexe A : Plan de l'IHM</a>	12
<a href="#">Annexe B : Logigramme des actions possibles sur les alignements avec l'interface</a>	13
<a href="#">Annexe C : Boutons présents dans l'interface</a>	14

## Objectifs

### ***Alignement du référentiel des personnes physiques de la Direction Juridique (DJ) avec celui de la Direction des Collections (DDCOL)***

#### **Objectif général**

L'objectif général est d'aligner deux jeux de données pour effectuer le rapprochement entre le système d'information juridique et celui documentaire. Ces jeux sont les suivants :

- les personnes physiques issues du système d'information juridique (DJ) :
  - un jeu avec les personnes ayant des contributions
  - un second jeu avec les personnes n'ayant pas de contributions
- les personnes physiques de la DDCOL

#### **Structure des jeux de données initiaux**

##### **Côté DJ**

Matricule
Prénom
Nom
Pseudo nom
Pseudo prénom
Civilité
Date de décès
Commentaire homonymie
Contribution
Nombre de contributions

Tableau 1 Champs disponibles côté DDCOL pour l'alignement

##### **Côté DDCOL**

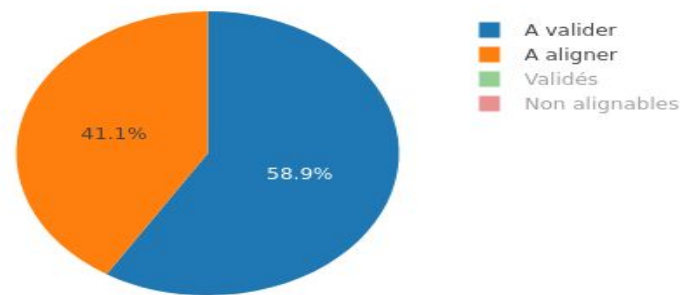
businessIdentifier
conceptIdentifier
Nom
Type de libellé
Note qualité
Sexe
Date de naissance
Date de décès
Type d'entité

Tableau 2 Champs disponibles côté DDCOL pour l'alignement

### ***Vérification des alignements réalisés avec Talend***

La première étape de l'alignement réalisée avec Talend a permis d'aligner 60% (42240) des matricules de la DJ (sur un total de 71780).

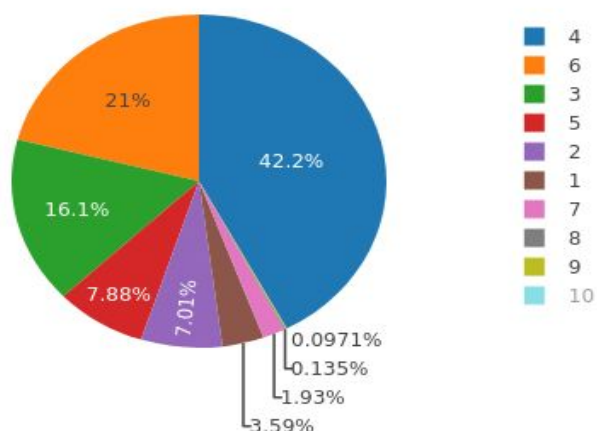
### Statut des matricules DJ dans leur alignement avec la DDCOL



**Figure 1 Statut des matricules DJ dans leur alignement avec les concepts de la DDCOL**

Cependant, les alignements réalisés sont pour la plupart peu fiables en raison des homonymes et du peu d'éléments qui les ont permis. Chaque alignement s'est par conséquent vu attribuer un indice de confiance entre 1 (le plus faible) et 9 (le plus fort) pour aider le métier dans la validation de ces alignements.

### Indice de confiance des alignements à vérifier ou vérifiés



**Figure 2 Indice de confiance des alignements réalisés avec Talend**

L'interface réalisée devra donc permettre de valider un à un les alignements réalisés avec Talend, de manière à rejeter toute erreur d'alignement, et à valider les alignements qui paraissent corrects.

### ***Création de nouveaux alignements***

En raison des différences de graphie de l'état civil d'un même personne dans la DJ et dans la DDCOL, certains alignements n'ont pas pu être effectués. L'interface doit donc permettre de réaliser ces alignements grâce à la recherche du bon concept dans les données de la DDCOL.

# Interface réalisée

## *Choix*

### Langage de programmation

Le choix du langage de programmation de l'interface n'étant pas important ici, j'ai fait le choix d'une interface réalisée avec **Python3**.

### Outils

L'interface étant une interface web, mon choix s'est porté sur le framework web Python **Flask**. Le stockage des données est réalisé dans une base de données relationnelle **PostgreSQL** dans sa version 11, avec laquelle l'interface communique par l'ORM **SQLAlchemy**. Un accès graphique à la base de données est possible via DBBeaver.

Packages Python requis pour le fonctionnement de l'interface :

```
certifi==2020.6.20
chardet==3.0.4
Click==7.0
click-conf==0.0.1
click-stream==0.0.10
Flask==1.1.2
Flask-SQLAlchemy==2.4.3
idna==2.10
itsdangerous==1.1.0
Jinja2==2.11.2
MarkupSafe==1.1.1
psycpg2==2.8.5
pyaml==20.4.0
PyMySQL==0.9.3
python-dotenv==0.13.0
PyYAML==5.3.1
retrying==1.3.3
six==1.15.0
SQLAlchemy==1.3.18
Unidecode==1.1.1
urllib3==1.25.9
Werkzeug==1.0.1
WTForms==2.3.1
```

## Structure des données

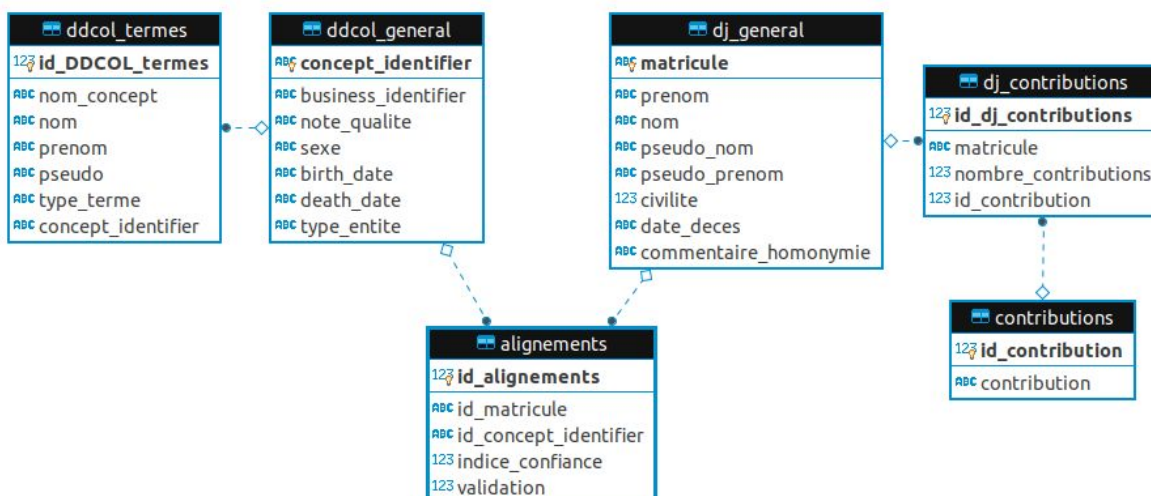


Figure 3 Modèle relationnel des données de l'interface

- La table « Alignements » est la table contenant l'ensemble des matricules DJ, pour lesquels un « id\_alignements » est attribué à chacun.

Ceux qui sont alignés avec la DDCOL ont l'attribut « id\_concept\_identifieur » non null et rempli.

En cas d'alignement, l'« indice de confiance » est compris entre 1 et 9 et la « validation » vaut 0.

En cas d'absence d'alignement, l'« indice de confiance » vaut 0 et la validation » vaut 0.

- Côté DDCOL :

- La table « ddcot\_general » contient les informations redondantes des concepts de la DDCOL. En effet, en raison des termes préférentiel et alternatif pour le nom d'un même concept, les informations d'identifiants (business et concept identifiants), de sexe, de dates de naissance et de mort, de fonction (contenue dans la note qualité) et de type d'entité sont répétées dans le jeu de données initial.

10234328	00f0013fdpqq	Elkaïm Bollinger, Renée	Terme alternatif en typo riche	Productrice radio. France	Femme		Personnes physiques
10234328	00f0013fdpqq	Elkaïm, Renée	Terme alternatif en typo riche	Productrice radio. France	Femme		Personnes physiques
10234328	00f0013fdpqq	Elkaïm Bollinger, Renée	Terme préférentiel en typo riche	Productrice radio. France	Femme		Personnes physiques

Figure 4 Exemple d'un même concept ayant plusieurs termes nom et prénom

- La table « ddcot\_termes » contient par conséquent les termes nom, prénom et pseudo du concept, ainsi que le type de terme dont il s'agit.

- Côté DJ :

- La table « dj\_general » contient les informations de chaque matricule, exceptées les contributions. En effet, un même matricule peut avoir plusieurs contributions.

70	ALBERT	UDERZO	UDERZO	ALBERT		Autres co-auteurs	3
70	ALBERT	UDERZO	UDERZO	ALBERT		Travaux Littéraires	27
70	ALBERT	UDERZO	UDERZO	ALBERT		Artiste dramatique	13
70	ALBERT	UDERZO	UDERZO	ALBERT		Oeuvre préexistante littéraire	63
70	ALBERT	UDERZO	UDERZO	ALBERT		Autres auteurs associés	9

**Figure 5 Exemple d'un même matricule ayant plusieurs contributions**

- La table « dj\_contributions » contient l'identifiant du label de la contribution, ainsi que le nombre de contributions de ce type pour chacun des matricules.
- Les labels des contributions étant normalisés, ils sont stockés dans la table « contributions » pour éviter une redondance des labels dans la table « dj\_contributions ».

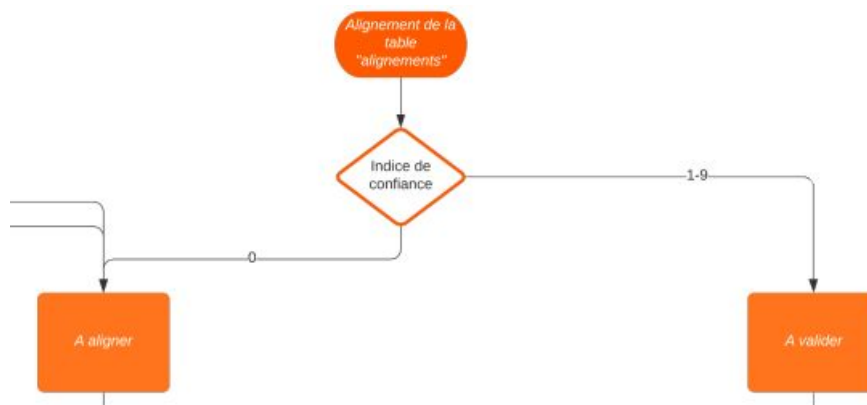
## ***Structure de l'interface***

### **Description générale de l'interface**

L'interface repose sur les alignements de la table « alignements » qui est la seule table dans laquelle l'application écrit ou modifie les données. Chaque matricule de la DJ a un « id\_alignements » qui lui est attribué ; en cas d'alignement, l'id du concept de la DDCOL est ajouté, ainsi que les indices de confiance et code de validation associés. Ces opérations sont effectuées au travers des différentes pages de l'interface.

L'interface (voir Annexe A : Plan de l'IHM) se divise en plusieurs parties, accessibles par la barre de navigation :

- la page d'accueil qui affiche deux graphiques de statistiques permettant de voir la volumétrie des données restant à traiter ou traitées
- deux pages permettant d'effectuer des actions ; ces deux pages sont distinguées par la valeur de l'indice de confiance de l'alignement (Figure 6.1 : Méthode de triage des alignements pour les pages de validation et d'alignement) :



**Figure 6 Méthode de triage des alignements pour les pages de validation et d'alignement**

- « /alignements\_a\_valider » permet de valider les alignements réalisés avec Talend
- « /alignements\_a\_aligner » permet de rechercher des personnes dans le jeu de la DDCOL, et d'aligner la personne trouvée avec le matricule affiché
- deux pages recensant les actions faites sur l'interface et les alignements, sans possibilité de remodifier ces alignements :
  - « /alignements\_valides » affiche 10 par 10 les alignements validés dans l'interface
  - « /alignements\_impossibles » affiche 10 par 10 les matricules déclarés dans l'interface comme impossibles à aligner à cause de l'absence d'un équivalent dans les données de la DDCOL

D'autres pages sont accessibles par rebond :

- la page « /matricule/<id\_matricule> » permet d'afficher un seul matricule ; des actions et redirections sont proposées en fonction du statut (à valider, à aligner, validé, non alignable) de son alignement
- la page « /concept/<id\_concept\_identifieur> » permet de voir un concept spécifique de la DDCOL, s'il apparaît dans la table des alignements ; comme pour la page « /matricule/<id\_matricule> », des actions sont possibles selon le statut de l'alignement

Deux URLs sont utilisées de manière invisible pour l'utilisateur en plus des précédentes :

- l'URL « /alignements\_a\_modifier » permet d'effacer en base l'alignement qui était proposé pour le matricule, et de rediriger l'utilisateur vers la page d'alignement de ce matricule
- l'URL « /searchbox » est uniquement appelée par le script JavaScript du template « partials/searchbox.html » ; elle permet de rechercher dans la base de données un maximum de 100 résultats correspondant aux lettres tapées dans la barre de recherche de la barre de navigation ; cette recherche est mise à jour à chaque caractère rentré.

## Description précise des URLs

### « /accueil »

L'URL « /accueil » ne prend pas de paramètres.

### « /alignements\_a\_valider »

Paramètre	Type	Par défaut	Commentaire
<b>id</b>	int		Identifiant de l'alignement dans la table «alignements»
<b>page</b>	int	1	Permet de naviguer dans les pages des alignements à valider
<b>validation</b>	str		Si « True », permet de valider l'alignement dont l'identifiant est le paramètre id ; si « False », supprime l'alignement

Tableau 3 Paramètres de l'URL "/alignements\_a\_valider"

## « /alignements\_a\_faire »

Paramètre	Type	Par défaut	Commentaire
concept_choisi	list		Ce paramètre récupère la valeur de la check-box HTML : cette valeur est composée ainsi : id_conceptIdentifiant@id_alignement
id	int		Identifiant de l'alignement dans la table «alignements»
libre	str		Chaîne de caractères (non sensible à la casse et à l'accentuation) contenue dans l'attribut « nom_concept » du concept de la DDCOL
nom	str		Chaîne de caractères strictement égale (mais non sensible à la casse et à l'accentuation) au nom du concept de la DDCOL
page	int	1	Permet de naviguer dans les pages des alignements à valider
prenom	str		Chaîne de caractères strictement égale (mais non sensible à la casse et à l'accentuation) au prénom du concept de la DDCOL
rechercher	bool	False	Si True, la recherche par nom, prénom ou texte libre est lancée dans le jeu de données de la DDCOL ; si False, elle ne l'est pas

Tableau 4 Paramètres de l'URL "/alignements\_a\_faire"

## « /alignements\_valides »

Paramètre	Type	Par défaut	Commentaire
page	int	10	Permet de naviguer dans les pages des alignements à valider

Tableau 5 Paramètres de l'URL "/alignements\_valides"

## « /alignements\_impossibles »

Paramètre	Type	Par défaut	Commentaire
id	int		Identifiant de l'alignement dans la table «alignements»
impossibilite	bool	False	Si True, l'alignement dont l'id est donné dans le paramètre « id » est passé en impossible à aligner (l'attribut « validation » de la table «alignements» est mis à 2)
page	int	10	Permet de naviguer dans les pages des alignements à valider

Tableau 6 Paramètres de l'URL "/alignements\_impossibles"



## « /matricule/<id\_matricule> »

Paramètre	Type	Par défaut	Commentaire
id_matricule	str		Identifiant du matricule dans la table «alignements»

Tableau 7 Paramètres de l'URL « /matricule/<id\_matricule> »

## « /concept/<id\_concept> »

Paramètre	Type	Par défaut	Commentaire
id_concept	str		Identifiant du concept dans la table « alignements »

Tableau 8 Paramètres de l'URL « /concept/<id\_concept> »

## « /alignements\_a\_modifier »

Paramètre	Type	Par défaut	Commentaire
id	int		Identifiant de l'alignement dans la table «alignements»

Tableau 9 Paramètres de l'URL « /alignements\_a\_modifier »

## « /searchbox »

Paramètre	Type	Par défaut	Commentaire
text	str		Chaîne de caractères tapée dans la barre de recherche générale de la barre de navigation

Tableau 10 Paramètres de l'URL « /searchbox »

## *Fonctionnalités*

Plusieurs actions sont possibles sur un alignement de la table « alignements » (Annexe B : Logigramme des actions possibles sur les alignements avec l'interface). Un alignement peut présenter deux cas de figure :

- si l'indice de confiance est à 0 et la validation à 0, alors l'alignement n'a pas eu d'alignement proposé suite à l'étape Talend ; il faut donc l'aligner si possible
- si l'indice de confiance est compris entre 1 et 9, alors un alignement est proposé et il faut le valider ou le rejeter

### **Validation d'un alignement**

La page « /alignements\_a\_valider » permet quatre actions :

- **Valider** l'alignement proposé (Figure 10 Bouton "Valider") : l'indice de confiance passe alors à 10, l'attribut validation à 1. Sur la page du matricule (« /matricule/<id\_matricule> »), plus aucune action n'est alors possible ; l'alignement est considéré comme définitif et sûr.
- **Rejeter** l'alignement proposé (Figure 11 Bouton "Rejeter") : l'indice de confiance passe à 0, l'attribut validation reste à 0, et l'« id\_concept\_identifier » qui était présent est supprimé. Cet alignement rejeté sera alors visible dans la page « /alignements\_a\_faire ».
- Alignement **non alignable** (Figure 12 Bouton "Non alignable") : l'indice de confiance passe à 0, et la validation à 2. Sur la page du matricule (« /matricule/<id\_matricule> »), plus aucune action n'est alors possible ; l'alignement impossible est considéré comme définitif et sûr.
- **Suivant** : permet de se diriger vers un alignement suivant. Aucune action n'est alors effectuée sur l'alignement actuellement proposé.

### Alignement à réaliser

La page « /alignements\_a\_faire » propose deux actions quand une recherche de concept DDCOL n'a pas encore été effectuée :

- Alignement **non alignable** (Figure 12 Bouton "Non alignable") : l'indice de confiance passe à 0, et la validation à 2. Sur la page du matricule (« /matricule/<id\_matricule> »), plus aucune action n'est alors possible ; l'alignement impossible est considéré comme définitif et sûr.
- **Suivant** : permet de se diriger vers un alignement suivant. Aucune action n'est alors effectuée sur l'alignement actuellement proposé.

Pour aligner le matricule DJ avec un concept DDCOL, il faut effectuer une recherche manuelle selon trois critères (ne remplir qu'un seul est suffisant) :

Figure 7 shows a search bar with four input fields and one button. The fields are labeled 'Nom de famille exact', 'Prénom exact', and 'Etat civil libre'. The button is labeled 'Rechercher' and is highlighted with a green border.

**Figure 7 Barre de recherche dans les concepts de la DDCOL**

- le champ « nom » recherche exactement le nom donné dans l'attribut « nom » de la table « ddcol\_termes ». La recherche n'est pas sensible à la casse et à l'accentuation.
- le champ « prénom » recherche exactement le prénom donné dans l'attribut « prenom » de la table « ddcol\_termes ». La recherche n'est pas sensible à la casse et à l'accentuation.
- la chaîne rentrée dans le champ « état civil libre » est recherchée dans l'attribut « nom\_concept » de la table « ddcol\_termes ». La recherche n'est pas sensible à la casse et à l'accentuation.

Si des résultats apparaissent, il est possible de cocher le concept avec lequel on souhaite aligner le matricule. Un bouton « **Enregistrer** » disponible en haut de page (Figure 14 Bouton "Enregistrer") permet d'enregistrer ce choix : l'indice de confiance passe à 5, la validation reste à 0, et l'id du concept choisi est inséré dans « id\_concept\_identifier ». Une redirection est effectuée vers la page du matricule « /matricule/<id\_alignement> ».

Sur cette page de redirection, plusieurs actions sont désormais possibles :

- **Valider** l'alignement proposé (Figure 10 Bouton "Valider") : l'indice de confiance passe alors à 10, l'attribut validation à 1. Sur la page du matricule (« /matricule/<id\_matricule> »), plus aucune action n'est alors possible ; l'alignement est considéré comme définitif et sûr.

- **Rejeter** l'alignement proposé (Figure 11 Bouton "Rejeter"): l'indice de confiance passe à 0, l'attribut validation reste à 0, et l'« id\_concept\_identifier » qui était présent est supprimé. Cet alignement rejeté sera alors visible dans la page « /alignements\_a\_faire ».
- **Modifier** l'alignement proposé (Figure 13 Bouton "Modifier") : l'indice de confiance passe à 0, l'attribut validation reste à 0, et l'« id\_concept\_identifier » qui était présent est supprimé. Cet alignement modifié sera alors visible dans la page « /alignements\_a\_faire ».
- **Suivant** : permet de se diriger vers un alignement suivant. Aucune action n'est alors effectuée sur l'alignement actuellement proposé.

## Recherche dans les jeux de la DJ et de la DDCOL

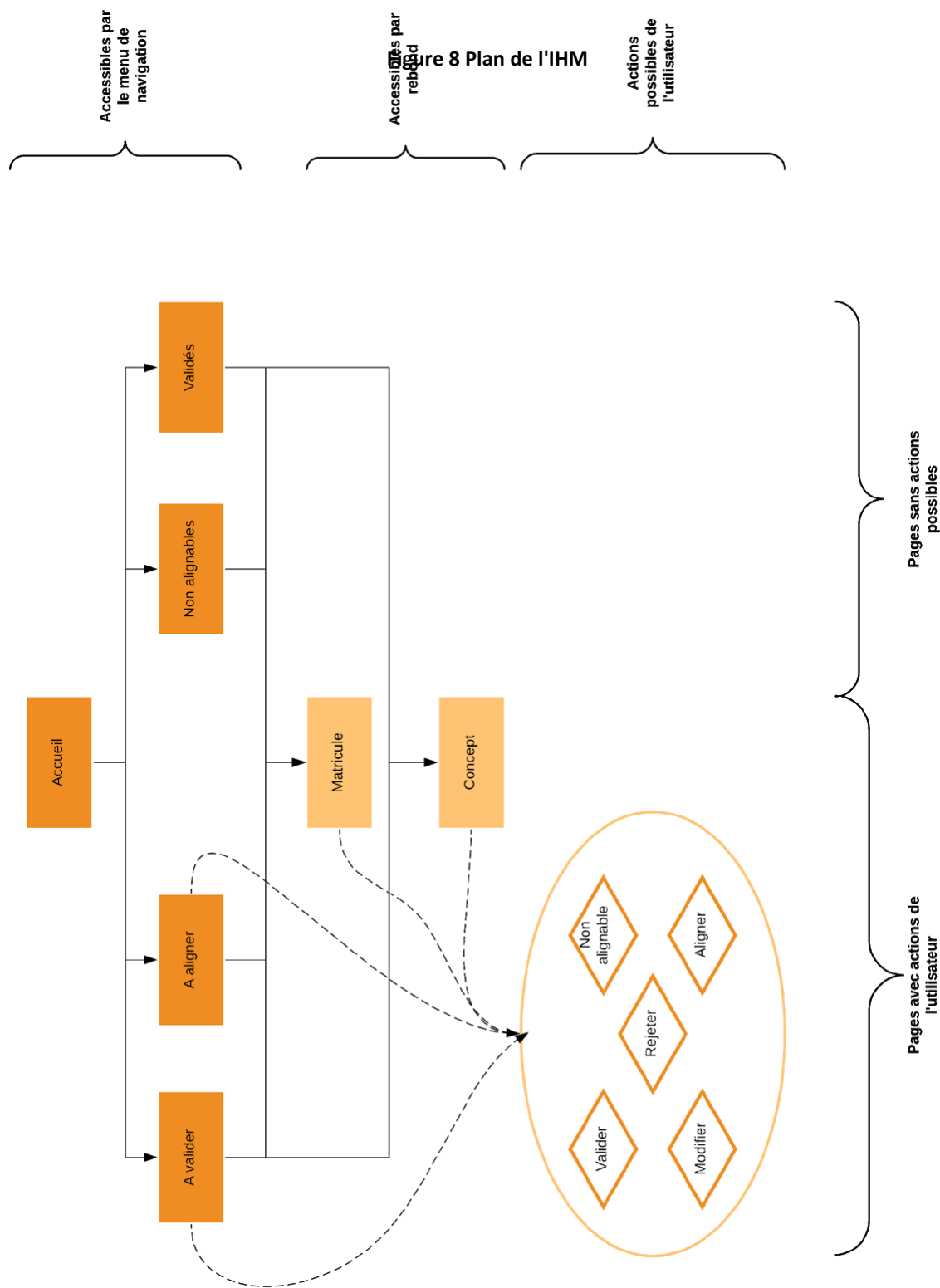
Dans la barre de recherche de la barre de navigation, il est possible de rechercher par :

- identifiant de matricule DJ
- identifiant de concept DDCOL
- nom de matricule ou concept
- prénom de matricule ou concept
- pseudos de matricule ou concept

La recherche se met à jour à chaque caractère rentré dans la barre de recherche. La recherche n'est pas sensible à la casse et aux accents. Une seule espace est autorisée dans la recherche

Annexes

Annexe A : Plan de l'IHM



## Annexe B : Logigramme des actions possibles sur les alignements avec l'interface

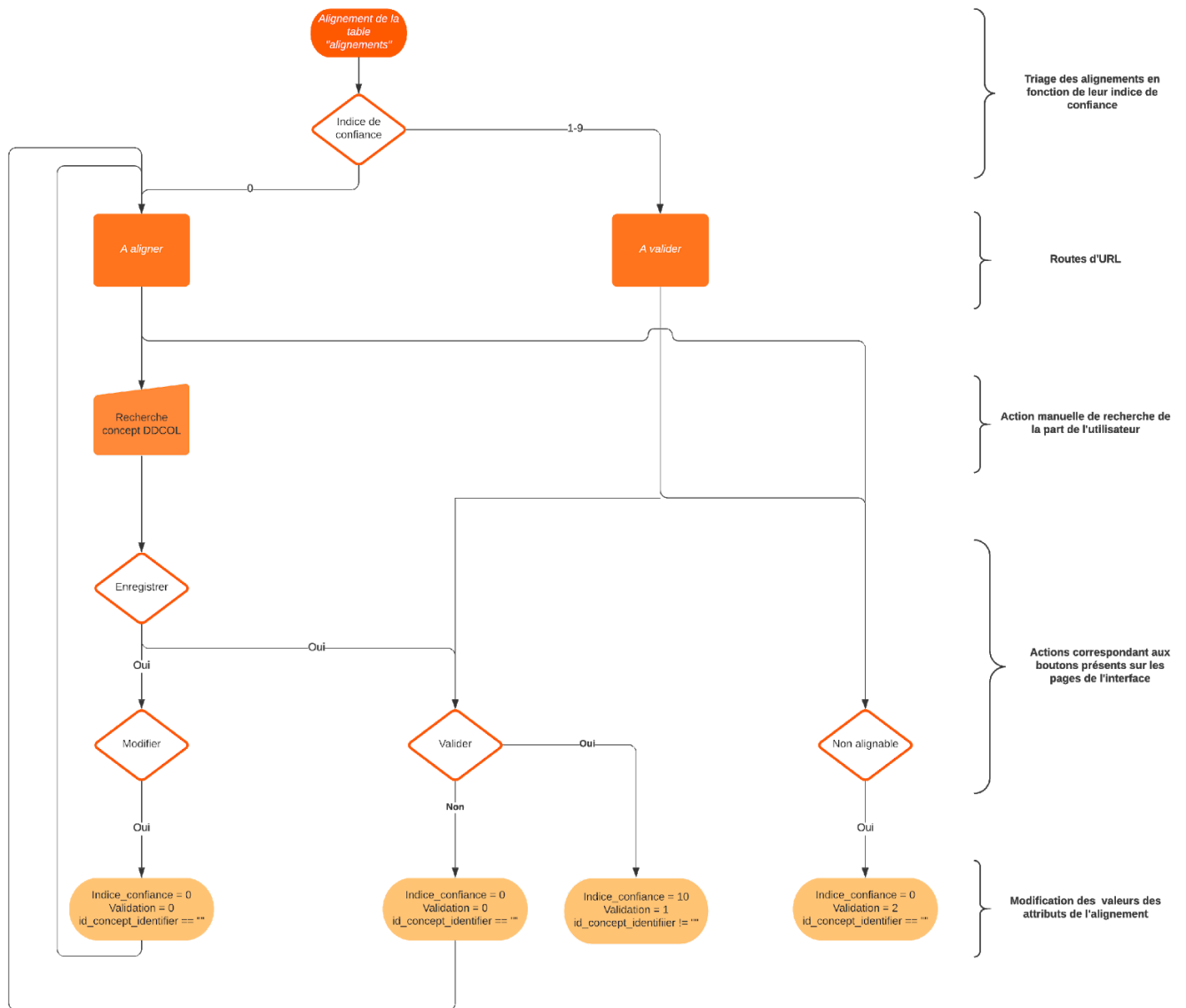


Figure 9 Logigramme des actions possibles sur les alignements avec l'interface

## Annexe C : Boutons présents dans l'interface

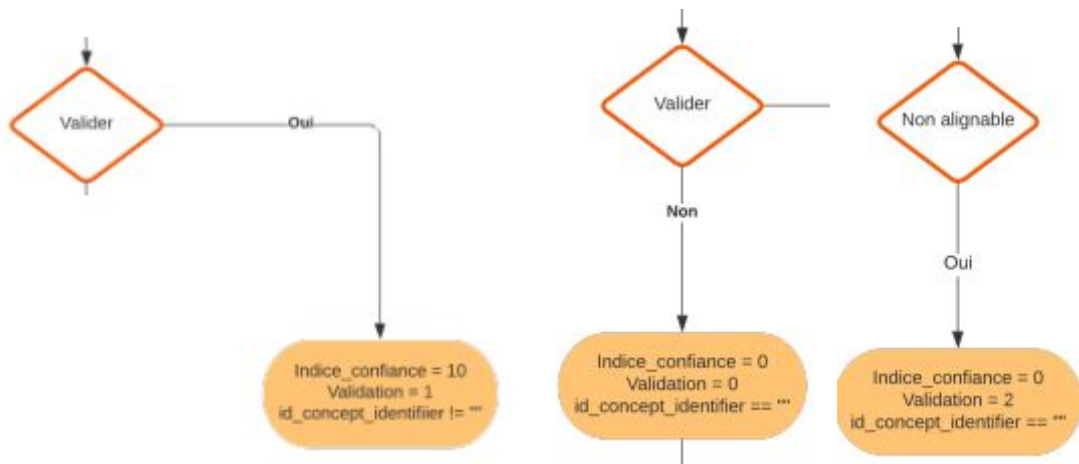


Figure 10 Bouton "Valider"

Figure 11 Bouton "Rejeter"

Figure 12 Bouton "Non alignable"

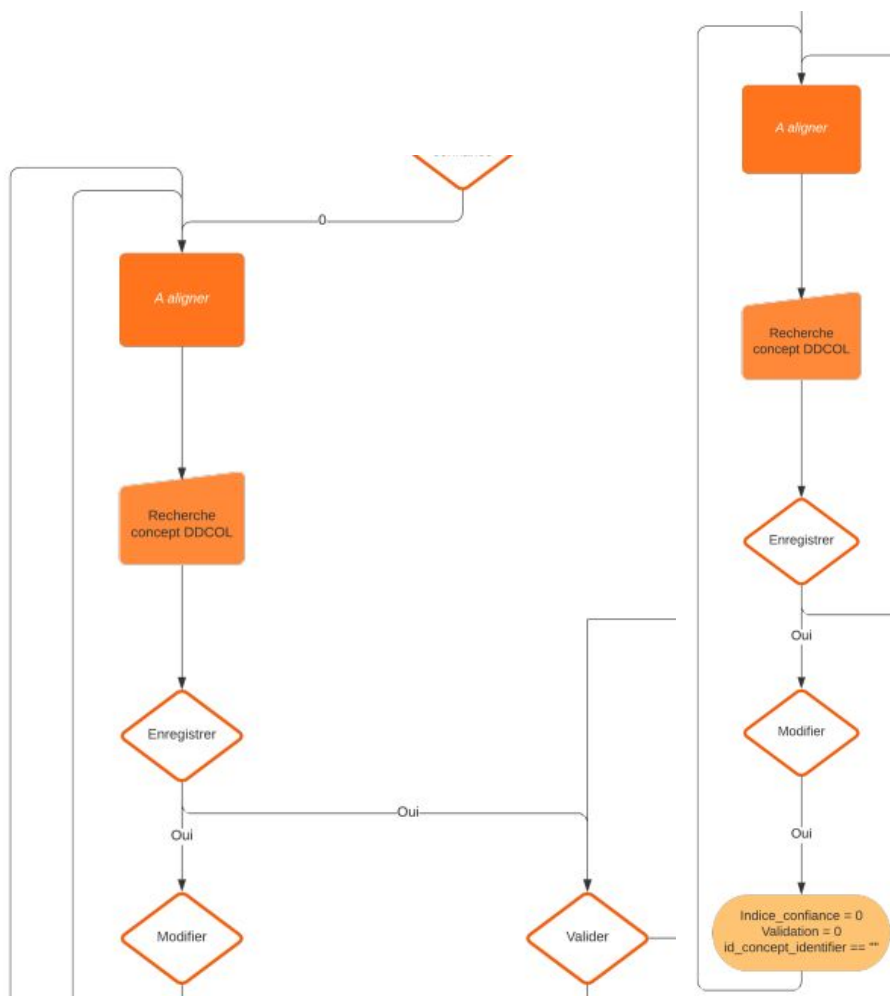


Figure 13 Bouton "Modifier"

Figure 14 Bouton "Enregistrer"