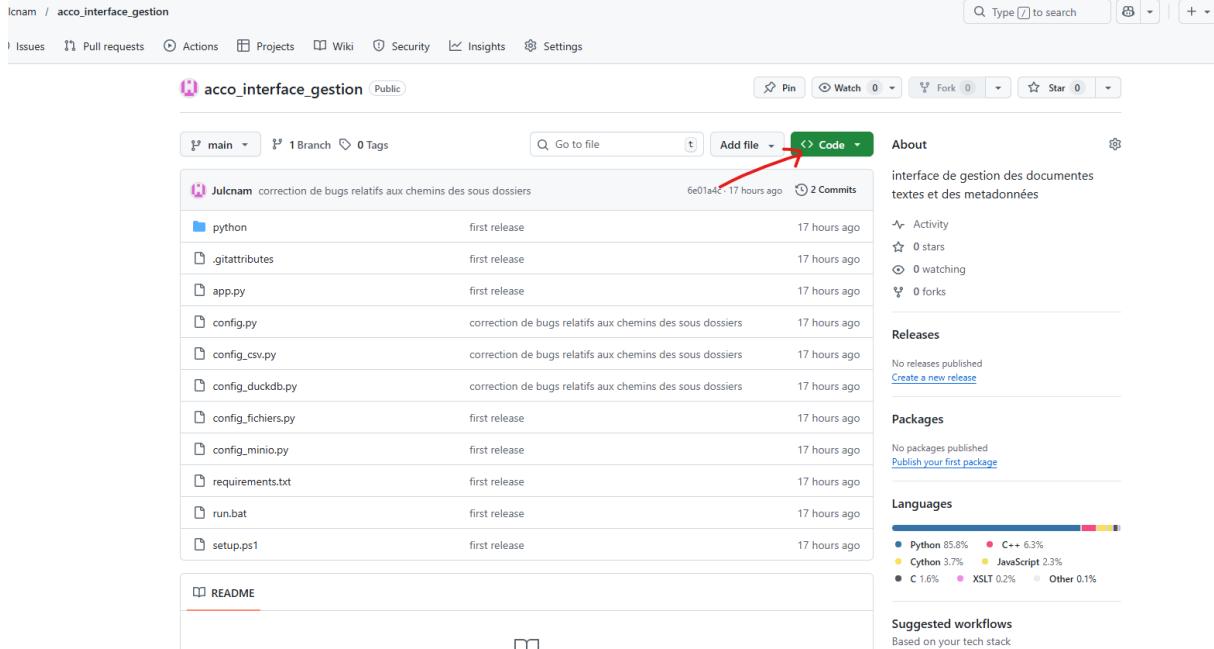


# Installation :

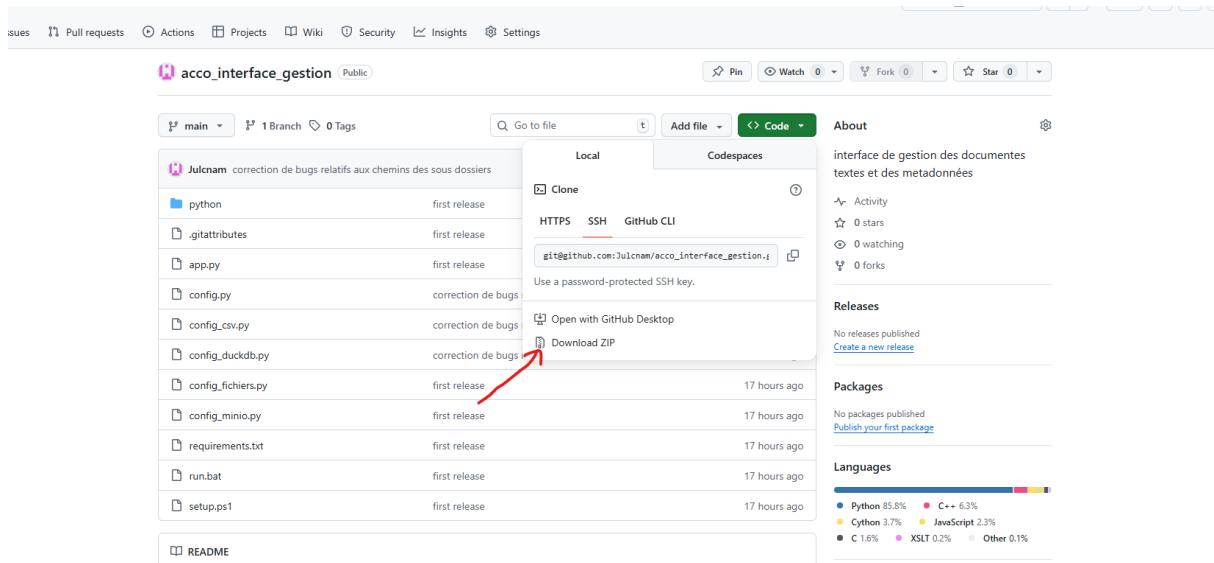
Rendez-vous sur GitHub avec ce lien :

[https://github.com/Julcnam/acco\\_interface\\_gestion](https://github.com/Julcnam/acco_interface_gestion)

Cliquez ensuite sur le bouton vert <>Code.



Et sur Download ZIP pour télécharger l'archive ZIP de l'interface.



Une fois l'archive téléchargée, vous pouvez l'extraire dans le dossier de votre choix.

## Lancement :

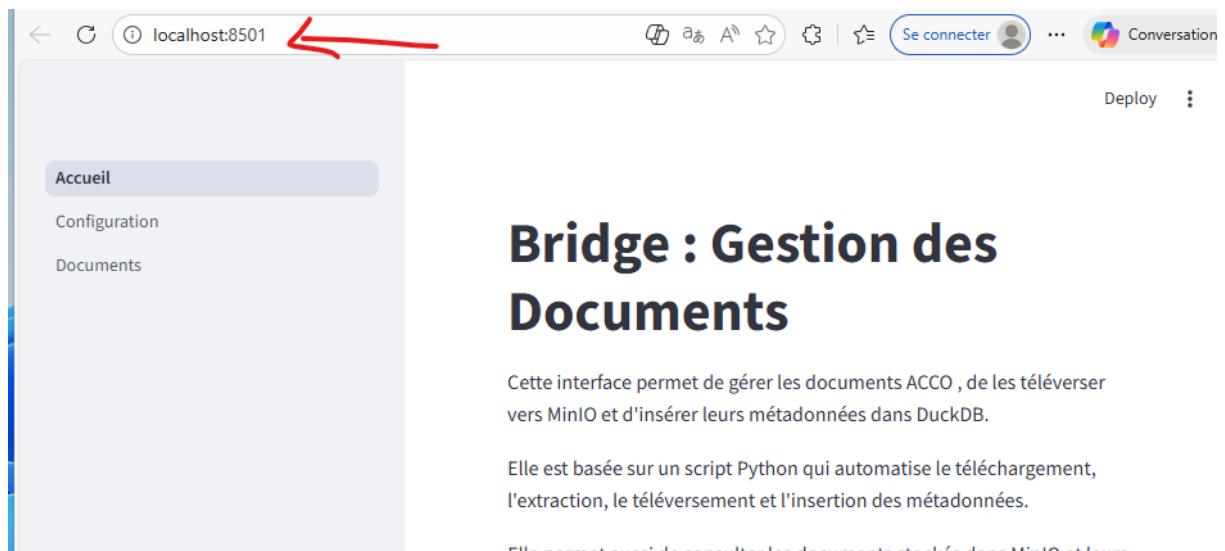
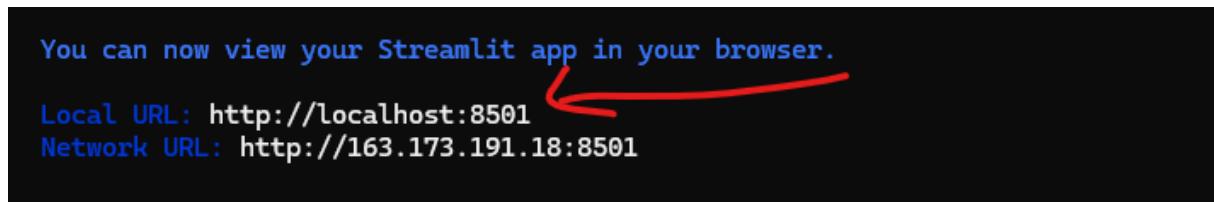
Pour lancer l'interface, il suffit de double-cliquer sur le fichier run.bat

📁 __pycache__	18/02/2026 17:23	Dossier de fichiers
📁 Bdd	18/02/2026 17:20	Dossier de fichiers
📁 Csv	18/02/2026 17:14	Dossier de fichiers
📁 python	18/02/2026 15:33	Dossier de fichiers
📁 Texts	16/02/2026 10:08	Dossier de fichiers
⚙️ .gitattributes	18/02/2026 16:48	Fichier source Git ... 1 Ko
⚙️ .gitignore	18/02/2026 16:43	Fichier source Git I... 1 Ko
🐍 app.py	18/02/2026 11:34	Fichier source Pyt... 5 Ko
🐍 config.py	18/02/2026 17:22	Fichier source Pyt... 2 Ko
🐍 config_csv.py	18/02/2026 17:23	Fichier source Pyt... 8 Ko
🐍 config_duckdb.py	18/02/2026 17:22	Fichier source Pyt... 4 Ko
🐍 config_fichiers.py	18/02/2026 17:22	Fichier source Pyt... 5 Ko
🐍 config_minio.py	18/02/2026 12:01	Fichier source Pyt... 3 Ko
📄 requirements.txt	17/02/2026 17:08	Document texte 1 Ko
⚙️ run.bat	16/02/2026 18:14	Fichier de comma... 1 Ko
🐍 setup.ps1	18/02/2026 15:32	Script Windows P... 1 Ko

La console va se lancer et effectuer des maj de python si besoin, à noter que l'interface embarque sa propre version de python portable pour éviter à l'utilisateur de devoir l'installer.

**IMPORTANT : Si c'est votre premier lancement, Streamlit peut vous demander un mail via la console, c'est purement optionnel et vous avez juste à appuyer sur entrée pour sauter cette étape.**

L'interface devrait s'ouvrir toute seule dans votre navigateur, sinon vous pouvez l'ouvrir manuellement via l'url locale fournie dans la console.



## Prise en main :

Les deux onglets principaux sont Scripts et Documents.

Dans un premier temps, Documents est inutile et va afficher une erreur, c'est normal, votre base de données métadonnées est encore vide.

Dans Scripts, vous avez des boutons permettant d'exécuter des scripts python.

## Configuration

Télécharger, extraire et convertir les fichiers .docx

Téléverser les fichiers vers MinIO onyxia

Insérer les métadonnées dans DuckDB

Créer le csv métadonnées

Chaque bouton exécute un Script python différent.

**IMPORTANT : une fois un bouton cliqué, cela ne sert à rien de cliquer sur un autre ou de cliquer plusieurs fois sur le même, vous pouvez suivre l'avancement du script sur la console.**

Le premier télécharge, extrait et converti les fichiers des accords d'entreprises dans un sous-dossier Texts à la racine des dossiers de l'interface.

📁 __pycache__	18/02/2026 17:23	Dossier de fichiers
📁 Bdd	18/02/2026 17:20	Dossier de fichiers
📁 Csv	18/02/2026 17:14	Dossier de fichiers
📁 python	18/02/2026 15:33	Dossier de fichiers
📁 Texts	16/02/2026 10:08	Dossier de fichiers
📄 .gitattributes	18/02/2026 16:48	Fichier source Git ...
📄 .gitignore	18/02/2026 16:43	Fichier source Git I...
📄 app.py	18/02/2026 11:34	Fichier source Pyt...
📄 config.py	18/02/2026 17:22	Fichier source Pyt...
📄 config_csv.py	18/02/2026 17:23	Fichier source Pyt...
📄 config_duckdb.py	18/02/2026 17:22	Fichier source Pyt...
📄 config_fichiers.py	18/02/2026 17:22	Fichier source Pyt...
📄 config_minio.py	18/02/2026 12:01	Fichier source Pyt...
📄 requirements.txt	17/02/2026 17:08	Document texte
📄 run.bat	16/02/2026 18:14	Fichier de comma...
📄 setup.ps1	18/02/2026 15:32	Script Windows P...

Le deuxième bouton, téléverse les fichiers qui ont été téléchargé sur le serveur s3 d'Onyxia permettant un accès réseau à ces documents. Ce script utilise les fichiers présents dans le sous dossier Texts, il faut donc garder les documents téléchargés dans Texts. **IMPORTANT : Le premier téléversement peut prendre très longtemps, environ 7 heures compte tenu du nombre important de fichiers.**

Le troisième bouton supprime tous les textes du serveur s3 d'Onyxia, il est à utiliser avec prudence, car une fois les fichiers supprimés, il faudra tout télécharger à nouveau.

Le quatrième bouton permet de constituer une base de données SQL locale avec les métadonnées de ces documents. A noter qu'une base de données SQL pré remplies est aussi disponible sur GitHub a ce lien :  
[https://github.com/Julcnam/acco\\_bdd](https://github.com/Julcnam/acco_bdd)

Le cinquième bouton créé un csv en damier à partir de la BDD SQL des métadonnées, ce csv est récupéré par Hugues pour faire fonctionner son interface de segmentation et d'annotation de textes. Un csv pré remplis est aussi disponible sur GitHub à ce lien :[https://github.com/Julcnam/acco\\_bdd](https://github.com/Julcnam/acco_bdd)

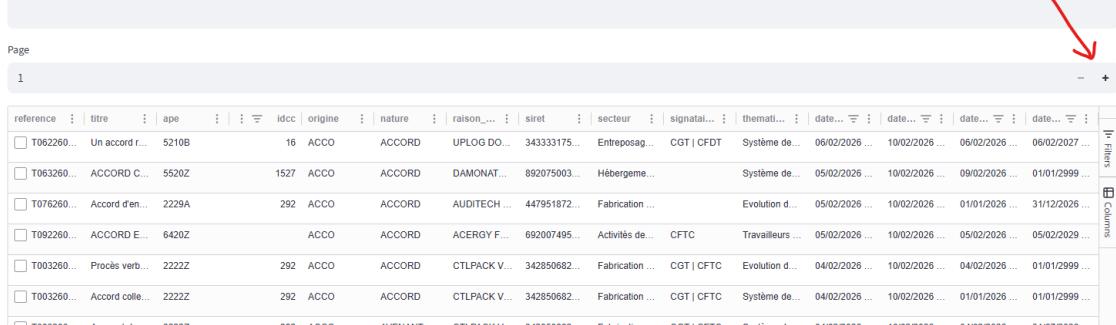
Une fois la base SQL des métadonnées constituée, il est possible d'accéder à l'onglet Documents. Cet onglet sert à inspecter l'état des données présentes dans la BDD, mais aussi à se renseigner sur un texte en particulier.

Chaque ligne représente un texte, on peut faire naviguer les textes suivant grâce au bouton + et revenir en arrière avec le bouton -.

## Documents

Liste des documents :

Recherche



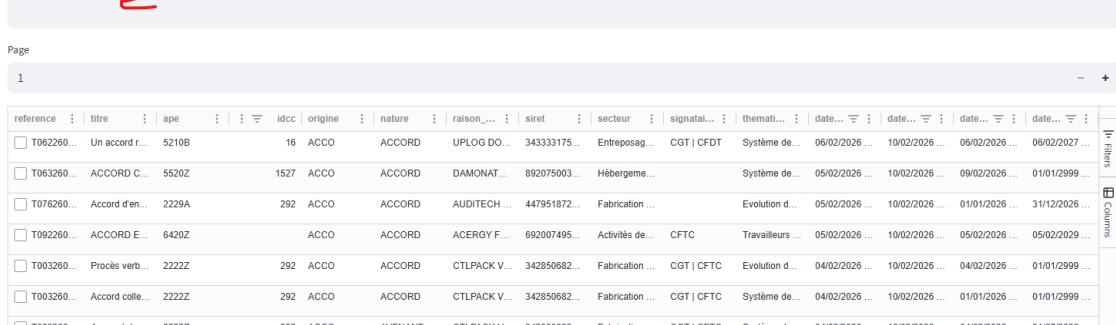
reference	titre	ape	idcc	origine	nature	raison_sociale	siret	secteur	signatair...	thematique	date...	date...	date...	date...	date...
T062260...	Un accord r...	5210B	16	ACCO	ACCORD	UPLOG DO...	343333175...	Entreposag...	CGT   CFDT	Système de...	06/02/2026 ...	10/02/2026 ...	06/02/2026 ...	06/02/2027 ...	
T063260...	ACCORD C...	5520Z	1527	ACCO	ACCORD	DAMONAT...	892075003...	Hébergeme...		Système de...	05/02/2026 ...	10/02/2026 ...	09/02/2026 ...	01/01/2999 ...	
T076260...	Accord d'en...	2229A	292	ACCO	ACCORD	AUDITECH...	447951872...	Fabrication ...		Evolution d...	05/02/2026 ...	10/02/2026 ...	01/01/2026 ...	31/12/2026 ...	
T092260...	ACCORD E...	6420Z		ACCO	ACCORD	ACERGY F...	692007495...	Activités de...	CFTC	Travailleurs ...	05/02/2026 ...	10/02/2026 ...	05/02/2026 ...	05/02/2029 ...	
T003260...	Procès verb...	2222Z	292	ACCO	ACCORD	CTLPACK V...	342850682...	Fabrication ...	CGT   CFTC	Evolution d...	04/02/2026 ...	10/02/2026 ...	04/02/2026 ...	01/01/2999 ...	
T003260...	Accord colla...	2222Z	292	ACCO	ACCORD	CTLPACK V...	342850682...	Fabrication ...	CGT   CFTC	Système de...	04/02/2026 ...	10/02/2026 ...	01/01/2026 ...	01/01/2999 ...	

Afin de faciliter la navigation, une barre de recherche permet d'effectuer une recherche sur la référence, le titre, l'entreprise, etc....

## Documents

Liste des documents :

Recherche



reference	titre	ape	idcc	origine	nature	raison_sociale	siret	secteur	signatair...	thematique	date...	date...	date...	date...	date...
T062260...	Un accord r...	5210B	16	ACCO	ACCORD	UPLOG DO...	343333175...	Entreposag...	CGT   CFDT	Système de...	06/02/2026 ...	10/02/2026 ...	06/02/2026 ...	06/02/2027 ...	
T063260...	ACCORD C...	5520Z	1527	ACCO	ACCORD	DAMONAT...	892075003...	Hébergeme...		Système de...	05/02/2026 ...	10/02/2026 ...	09/02/2026 ...	01/01/2999 ...	
T076260...	Accord d'en...	2229A	292	ACCO	ACCORD	AUDITECH...	447951872...	Fabrication ...		Evolution d...	05/02/2026 ...	10/02/2026 ...	01/01/2026 ...	31/12/2026 ...	
T092260...	ACCORD E...	6420Z		ACCO	ACCORD	ACERGY F...	692007495...	Activités de...	CFTC	Travailleurs ...	05/02/2026 ...	10/02/2026 ...	05/02/2026 ...	05/02/2029 ...	
T003260...	Procès verb...	2222Z	292	ACCO	ACCORD	CTLPACK V...	342850682...	Fabrication ...	CGT   CFTC	Evolution d...	04/02/2026 ...	10/02/2026 ...	04/02/2026 ...	01/01/2999 ...	
T003260...	Accord colla...	2222Z	292	ACCO	ACCORD	CTLPACK V...	342850682...	Fabrication ...	CGT   CFTC	Système de...	04/02/2026 ...	10/02/2026 ...	01/01/2026 ...	01/01/2999 ...	