# Étape 1 – Extraction des titres Wikipedia (Backlinks)

Date de génération : 2025-05-28 22:16:14

## 📦 Architecture modulaire

- BatchProcessingTitresExtraction hérite de BatchProcessing

- BatchWriterJSON gère la sérialisation, le découpage en lots, le flush automatique

## 🔁 Traitement

- Requête centralisée via requeteWikiMedia(url, params)

• Timeout 10s, retry sur exception

• Log d’erreur uniquement si exception ou lenteur (>1s)

- Traitement de milliers de backlinks

- Flush automatique toutes les 1000 lignes

- Log de performance ajouté : durée d'insertion par lot

## 🧹 Nettoyage effectué

- Méthode sauvegarder\_resultats() supprimée des classes de base

- La création de LigneProcess est centralisée dans traiter\_ligne()

- Requête API et logique d’erreur déplacées vers une fonction dédiée

## 🗂 Fichiers générés

- Fichiers JSON générés par lots de 1000 lignes

- Nommage typé : JD01\_step1\_YYYYMMDD\_HHMMSS\_batchX.json

- Rotation des logs automatique (10 Mo max, 4 backups)

- Aucun dépassement mémoire ou fuite détectée

Étape validée. Prêt pour l'étape 2 : enrichissement via SPARQL.