

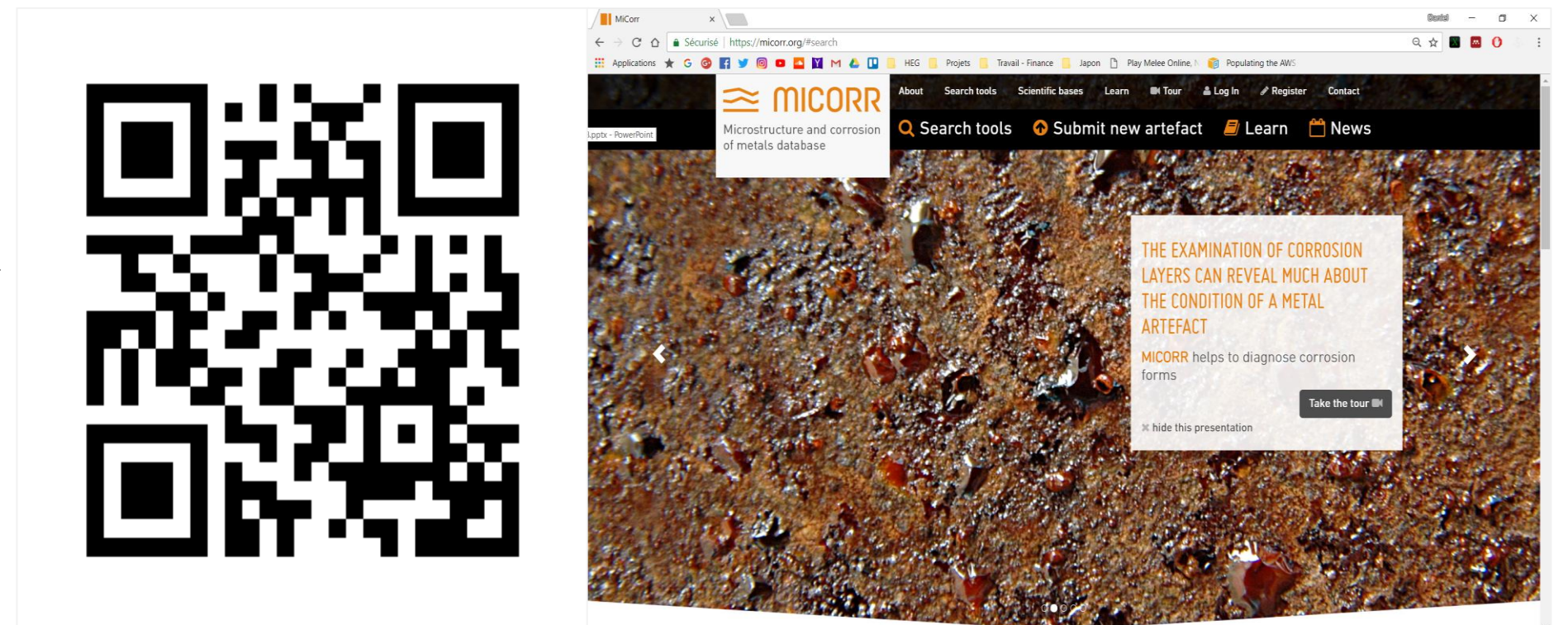
MiCorr – Indexation de documents

Daniel Chappuis

Contexte

MiCorr est une application de gestion destinée aux chercheurs dans le domaine de la restauration et de la conservation des artefacts métalliques. Elle permet aux chercheurs de comparer leurs résultats avec ceux présents dans la base de données de MiCorr.

On souhaite étendre cette recherche aux autres articles de la littérature.



<https://micorr.org/>

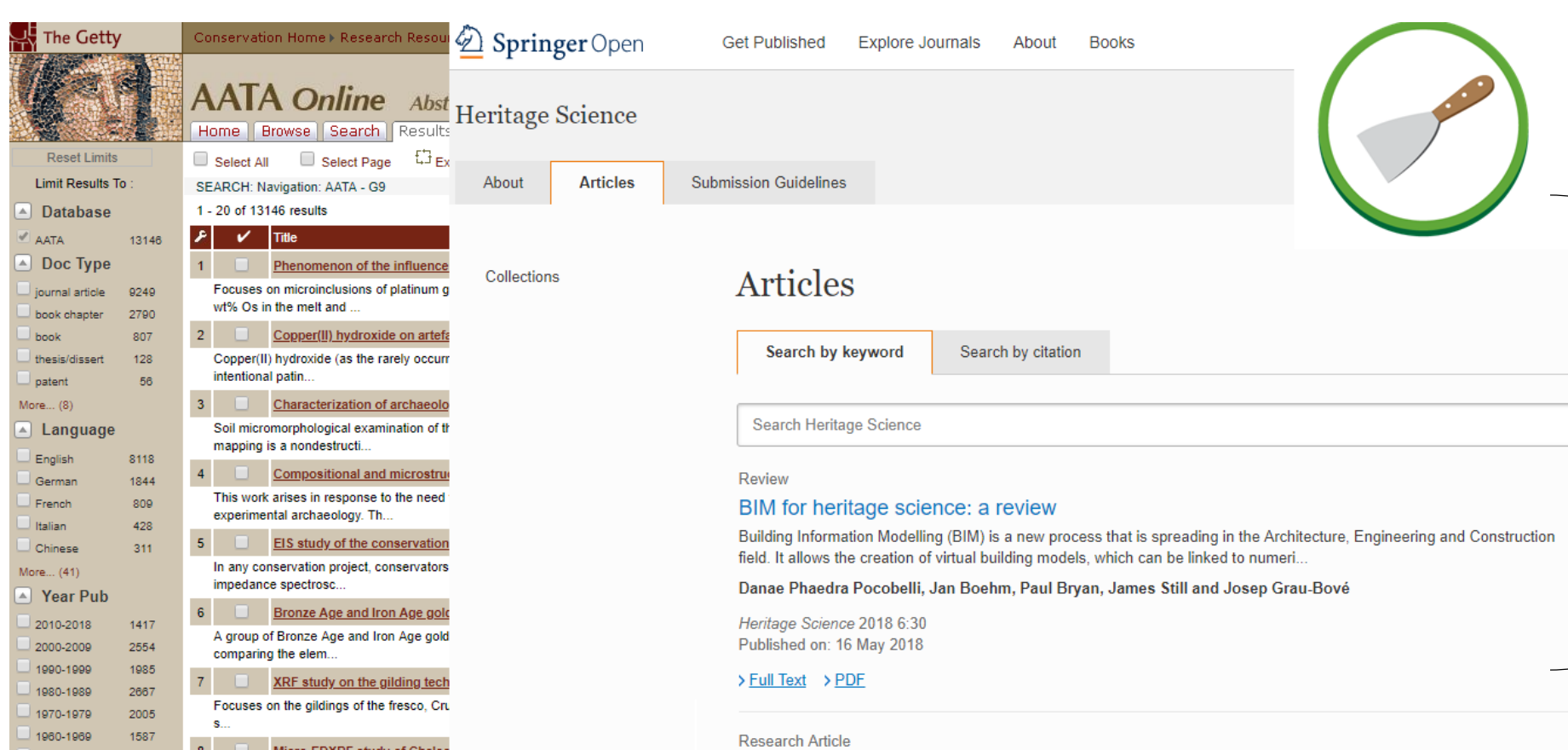
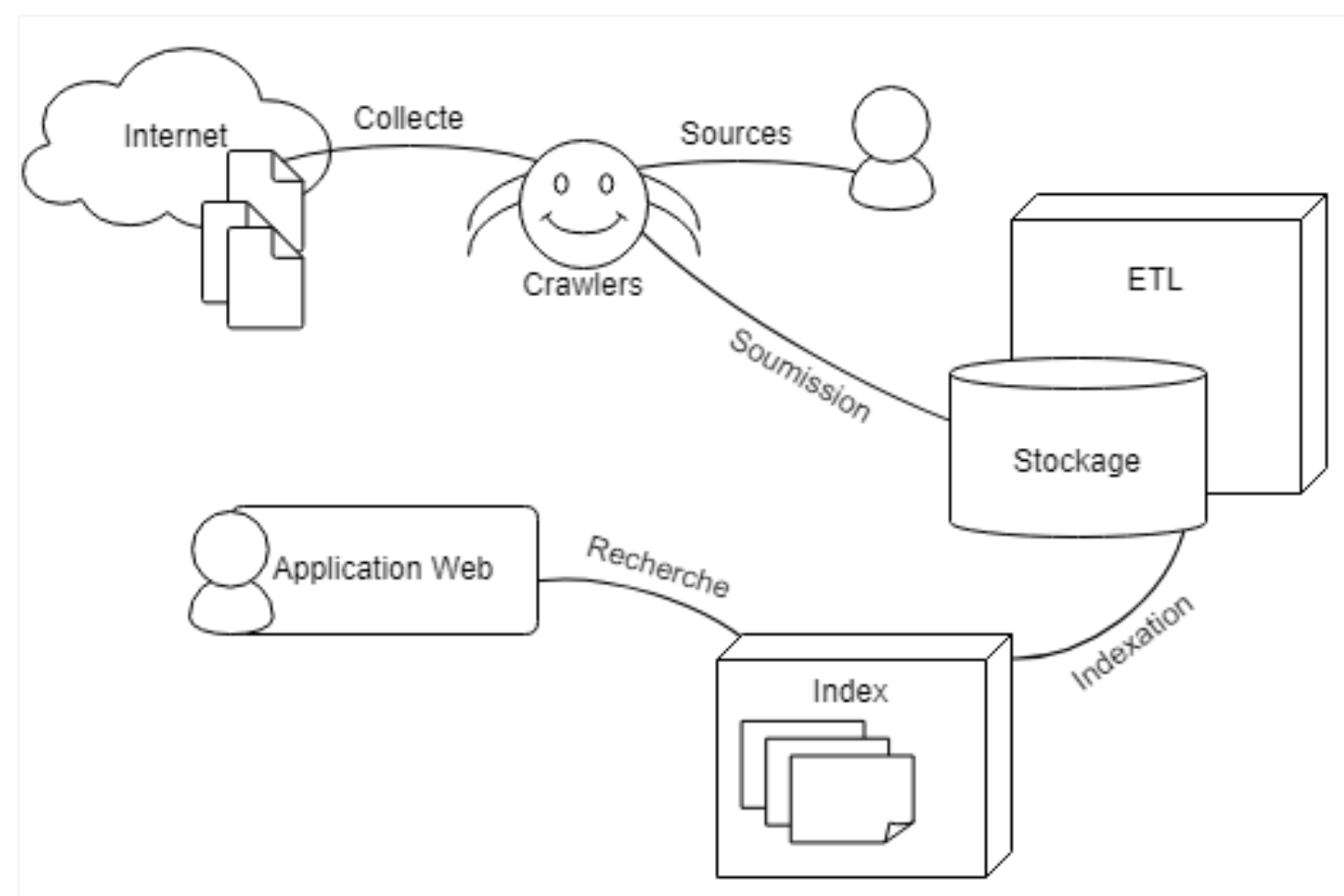
Problématique et objectif

Extraire des articles sur le web concernant un sujet précis et les structurer afin de les indexer. Valoriser les documents et enrichir les fonction de recherche avec les termes métier et étendre la fonctionnalité de recherche de l'application.

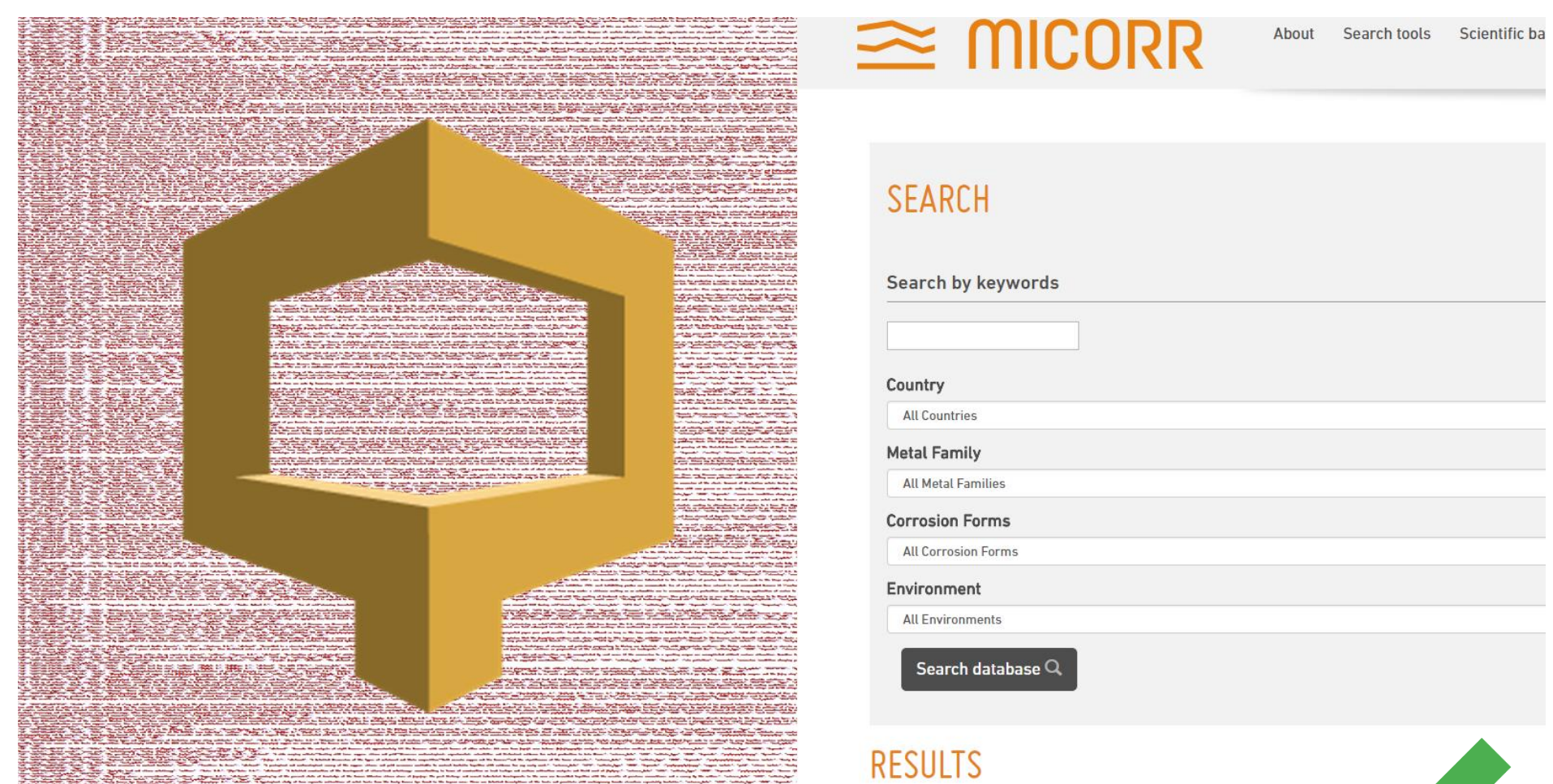
Approche

1. Récupérer les articles,
2. Les analyser et créer des documents,
3. Indexer ces documents,
4. Offrir une application de recherche.

La difficulté majeure est de retranscrire le jugement humain de la pertinence dans l'index.



Scrapy



Bénéfices et perspectives

- Contrôle des indicateurs de pertinence,
- Flexibilité de l'index,
- Personnalisation des requêtes,
- Suggestion,
- ...

