

Construction d'une base de données d'oeuvres culturelles en espéranto

Fabien Duchateau (prénom.nom@univ-lyon1.fr)

Contexte. L'[espéranto](#) est une langue internationale construite, parlée par plusieurs millions de personnes à travers le monde. Comme cette langue n'est rattachée à aucun état, l'une de ses critiques porte sur sa limitation au niveau culturel. Pourtant, de nombreuses oeuvres sont disponibles en espéranto, qui peuvent être soit des traductions d'oeuvres existantes (e.g., *la eta princo* pour *le petit prince*) soit des oeuvres originales (e.g., [Mr. Tot acetas mil okulojn](#)). À ce jour, il n'existe pas de liste la plus exhaustive possible rassemblant ces oeuvres espérantistes.

Objectifs. Ce projet a pour objectif le développement d'une base de données recensant un maximum d'oeuvres en espéranto (livres, poèmes, musique, films et vidéos, presse, radio, sites web, etc.). Ces oeuvres seront intégrées à partir de différentes sources de données. Pour enrichir les données intégrées, les oeuvres et auteur.e.s seront reliées à d'autres bases de connaissances (processus d'*entity linking*), par exemple pour avoir davantage d'informations sur l'auteur.e ou pour découvrir les versions d'une oeuvre en langue originale. Un site web pourra être développé afin de rechercher et visualiser les oeuvres culturelles espérantistes.

Organisation. Le travail du POM sera découpé ainsi :

- Modélisation de la base de données;
- Implémentation d'algorithmes d'intégration pour agréger des oeuvres issues de sources externes ([UEA katalogo](#), [BNF](#), [ANL](#), [Open Library](#), etc.) dans la base de données;
- Implémentation d'un algorithme pour la détection de liens vers d'autres référentiels ([VIAF](#), [WorldCat](#), [DBpedia](#), [Wikidata](#), [BNF](#), etc.);
- Développement d'un site web pour visualiser les données.

Contraintes. Base de données relationnelle, site web en HTML/CSS/PHP, prise en compte du multilinguisme.