

Transformer un fichier CSV en thésaurus SKOS

Cette semaine le PSIR vous propose une méthode simple et rapide pour transformer un fichier CSV ou une liste de mots clés en un thésaurus normalisé. CSV signifie (Comma-separated values = données séparées par des virgules) mais la règle est plus souple maintenant, un CSV peut aussi être séparé par (**un point-virgule ;**) ou (**tabulation \t**) voir d'autres caractères spécifiques.

Nous manipulons tous les jours des projets de type bases de données (classique ou du terrain), des sites web, des données d'archives ou d'images...

Ces projets utilisent souvent des mots clés pour indexer le contenu (image, objet, texte, enregistrement...).

Les formats de départ :

Les mots clés sont souvent gérés sous différentes formes :

- Une liste à plat (un mot clé par ligne)
- Une liste sur deux colonnes, la première pour les catégories, la seconde pour les mots clés
- Une liste plus complète avec une hiérarchie sur plusieurs colonnes. Cette représentation permet déjà de distinguer les mots clés du plus générique au plus spécifique.

Les inconvénients :

Les listes à plat privées ne servent en générale que pour un projet en particulier. Pour les partager avec d'autres projets, il faut disposer d'identifiants pérennes et d'un système de gestion centralisé pour contrôler et distribuer les index.

Voici la solution la plus simple pour transformer vos listes CSV en thésaurus au format SKOS normalisé :

- Nous disposons à la MOM d'une [plateforme centralisée](#) qui gère et héberge des thésaurus, elle est basée sur le logiciel [Opentheso](#).
- Opentheso dispose d'un module d'import de fichiers CSV, il accepte deux formats :
 - o Un fichier complet : qui peut être transformé directement en thésaurus avec des catégories/domaines et avec une hiérarchie à plusieurs niveaux (du plus générique au plus spécifique)
 - o Un fichier avec une seule colonne : qui contient une liste de mots clés

Après l'import, vos listes deviennent des thésaurus normalisés disposant d'identifiants pérennes de type ARK. Vous pouvez ensuite les utiliser à distance pour l'indexation, les partager, les exporter aux formats SKOS, JSON, Turtle, PDF et CSV.

[Pour plus d'informations sur comment importer un thésaurus en CSV](#)

NIVEAU DE
DIFFICULTÉ



Thésaurus

PERSONNES
CONCERNÉES



PDF

MINUTE_PSIR_#23