BOISDON Hugues

DERVILLE Thomas

Rapport Projet Datamining

But du projet :

L’objectif du projet était de créer un système de recommandations de monuments historiques. A partir d’une base de données composés d’une centaine de monuments et à l’aide d’une application récupérant des images « likées » par l’utilisateur, le système a pour but de recommander d’autres monuments qui correspondent à ses préférences.

Source des données et licences:

Les données des monuments proviennent toutes de Wikidata et ont été obtenues à l’aide du service de requête intégré. Toutes les images sont licenciées sous le CC BY-SA 4.0.

Les données concernant le style architectural et le pays ont été obtenues via l’API WikiBaseIntegrator qui permet de récupérer diverses données liées à un identifiant WikiData.

Taille des données :

Les données sont stockés sous deux formats différents : un fichier data au format JSON faisant environ 20 Ko et un dossier images contenant les images liées aux monuments de la base de données, dossier faisant environ … Mo.

Informations stockées pour chaque image :

label : Contient le nom (en anglais afin d’avoir le moins de valeurs inconnues) du monument.

link : Contient le lien wikimedia vers l’image principale du monument (image qui apparaît sur la page wikipédia du monument en général)

entityID : Contient l’identifiant WikiData du monument.

format : Contient le format de l’image principale (jpeg, png,…)

path : Contient le chemin relatif vers l’image contenu dans le dossier image stocké localement.

size : Contient la taille de l’image sous forme d’une liste à deux membres sous le format [ X, Y ]

sizeCategory :

tags : Contient une liste possédant les différents tags liés au monument entre autres le siècle, le pays et le/les mouvement artistiques liées au monument. Principalement utilisé dans le système de recommandation afin de différencier les différents monuments.

orientation :

artist : Contient le nom de l’auteur de la photo.

dateTaken : Contient la date à laquelle la photo a été prise.

appareil : Contient le nom de l’appareil avec lequel la photo a été prise.

dColors :

dProportions :

Informations concernant les préférences de l’utilisateur :

Modèles d’exploration de données :

Apprentissage machine :

Auto-évaluation :

Annotation des données :