

Data'Clear

February 12, 2021

Intégrer des données provenant de différentes sources en présence d'une ontologie (lire : un modèle conceptuel des données en UML) afin de préparer les données à l'apprentissage et la prédiction.

1 Les étapes

1.1 Chargement des données

La première étape consiste à charger les données des fichiers dans des bases de données adaptées. Nous optons pour deux systèmes, si possible différents, de base de données SQL pour les fichiers CSV ainsi qu'une base de données NoSQL pour le fichier JSON.

1.2 Définir le modèle ontologique

La seconde étape consiste à se mettre à la place des experts du domaine afin de définir un modèle ontologique (modèle conceptuel UML) permettant de décrire les données de façon unique.

1.3 Intégrer les données

La troisième étape consiste à intégrer les données des différentes sources dans le modèle ontologique (modèle conceptuel UML) en mettant en place un système de mappings de la forme

Requête sur la source $-i$ Données compatibles avec le format ontologique.

1.4 Graal

La quatrième étape consiste à utiliser Graal afin d'appliquer les mappings et les règles afin d'obtenir des données uniformes et complètes.

1.5 Apprentissage Prédiction / Visualisation

La cinquième étape consiste à déterminer qui va survivre au naufrage.