

Module de détection de similarité entre images

Contexte

Kernix développe [DBSE](#) au sein de son Lab une Data Management Platform [DMP], dont les objectifs sont d'une part d'agréger et d'enrichir des données multi-sources (données détenues par l'utilisateur et données extraites du web) et d'autre part d'exploiter ces données à des fins d'analyse prédictive et de recommandation. L'architecture de DBSE repose sur un noyau centré sur l'exploitation de deux bases de données NoSQL : MongoDB pour la gestion de documents non structurés et Neo4j pour le stockage de données relationnelles sous forme de graphe. Ce noyau est connecté à des modules de traitement de données spécifiques utilisant les outils du machine-learning et du traitement automatique du langage naturel [NLP] par exemple. Le Kernix Lab cherche continuellement à améliorer les modules préexistants et à intégrer de nouveaux modules à sa plateforme.

Objectifs du projet

L'objectif de ce projet IMPACT est de développer un module d'analyse d'images afin d'enrichir l'analyse sémantique des documents présents dans les bases de données par une analyse de similarité des images les accompagnant. Ce module permettra l'amélioration des recommandations générées par DBSE en ajoutant une nouvelle dimension d'analyse des données.

Stratégie

Plusieurs algorithmes existent pour comparer des images. Dans un premier temps, il s'agira d'implémenter un algorithme de détection de similarité basé sur l'histogramme de différentes composantes des images et un algorithme de keypoint matching (par exemple SIFT). Ces deux approches seront ensuite comparées et leur impact sur les performances de recommandation sera analysé.

Si nous disposons d'un temps suffisant ensuite, le candidat pourra mettre en place un algorithme fondé sur les réseaux de neurones (deep learning) afin de pouvoir analyser la similarité des images sur la base de leur contenu, plutôt que sur la base de leurs caractéristiques purement visuelles.

Contacts

François-Xavier Bois, directeur du Kernix Lab : fxbois@kernix.com

Imen El Karoui, Data scientist : ielkaroui@kernix.com

Joseph Pellegrino, Data scientist : jpellegrino@kernix.com