

Projet de fil rouge

Etudiants : Bernd STADELMAYER , Patrick CHEVARIER , ...

Sujet

Élaborer un programme de scraping qui à partir d'un ou plusieurs articles scientifiques , parcourt récursivement le chaînage des références bibliographiques et les mots clefs pour établir une banque des articles connexes.

A partir de bibliothèques existantes, établir un programme de machine learning qui à partir d'une classification (intéressant / pas intéressant) permet de classer les articles retrouvés en fonction de leur intérêt potentiel. Échantillonnage, classification, validation.

Utiliser les programmes réalisés sur le sujet du Machine Learning dans le domaine de la santé.

A partir des articles retrouvés, et d'un dictionnaire de pathologie (CIM10 ou autre) ; extraire de chaque article, les pathologies cibles et la taille des échantillons utilisés. Définir et utiliser une structure de donnée noSQL pour mémoriser les informations retrouvées. Utiliser un ETL pour homogénéiser les données extraites et les dictionnaires utilisés

Élaborer des tableaux de bord quantitatifs (mots déterminants, pathologies, domaines, pays, ...) sur le ML dans le domaine de la santé.

Le projet sera géré en suivant la méthode Agile, avec alternance des rôles tous les 2 jours.

Bonus possible : Impact des mots déterminants identifiés dans les réseaux sociaux.

Techno : Python, ML, MongoDB, Talend, Tableau ou PowerBI