

Analyse automatique de la difficulté de textes littéraires

Daive Baffa

Contexte

Dans le cadre du projet "Nous Lisons", l'objectif initial était de développer un outil basé sur les techniques de traitement automatique du langage. Ce projet avait pour mission d'évaluer la difficulté d'un texte dans différentes langues. Il offre ainsi un outil novateur pour les enseignants et accompagnateurs dans le domaine linguistique, permettant ainsi l'accompagnement et l'aide à des personnes ayant des difficultés, des troubles de l'apprentissage ou simplement des apprenants de langues étrangères.

Objectif

Sur la base des premiers travaux réalisés par Monsieur Bordet, l'objectif était de mettre en place une méthode permettant la création d'une liste de mots courants à partir d'un corpus de texte défini.

L'objectif étant de créer une application fournissant une analyse de la difficulté d'un texte littéraire, basé sur cette liste. L'application permettra d'afficher l'âge auquel un texte est accessible ainsi que le niveau CECR (allant de A1 à C2) requis.

En plus de cela, elle fournira une définition ainsi qu'une traduction pour chaque mot non courant. L'application devra supporter les langues française, espagnole et anglaise.

À terme toutes les langues officielles selon l'ONU devront pouvoir être intégrées, d'où l'importance de créer une méthode générique et évolutive

Méthodologie



Analyse de l'existant



État de l'art des technologies de NLP



Création d'une liste de mots les plus courants



Extraction des définitions et traductions pour la création d'une base de données



Application pour l'analyse d'un texte

Technologies et langage

