

Projet LINFO1104 - Orékhoff Alexandre ; L'Hermite Dumont Issambre

Extension - Généralisation de la formule du N-gramme **

Nous avons décidé de réaliser cette extension car celle-ci représentait, pour nous, le réel défi de ce projet. Nous l'avons implémenté en considérant un traitement de fichier à chaque prédiction. A chaque prédiction, nous considérons l'entrée et nous traitons tous les fichiers en cherchant cette dernière. Lorsqu'une occurrence est trouvée, le mot suivant celle-ci gagne un point.

En effet, les avantages sont multiples :

- Pas de chargement initial relativement long (jusqu'à 2 minutes pour certains projets)
- Construction d'un petit arbre et donc léger en terme de mémoire
- Arbre simple construit à chaque prédiction
- Temps de prédiction court : entre 5 et 10 secondes

Nous pensons que notre modèle est qualifiable "d'optimisé" et peut prétendre également à l'extension Extension - Optimisation de la formule du N-gramme * * *. Néanmoins, comme nous n'en étions pas sûr, nous avons réalisé d'autres extensions afin d'être sûr de respecter les consignes.

Extension - Optimisation de la formule du N-gramme * * *

Nous pensons que notre modèle est "optimisé". En effet, comme nous effectuons le parsing des fichiers lors de l'appel à la fonction "Press" nous connaissons déjà les mots à chercher dans les fichiers. Nous n'avons donc aucun temps d'attente au lancement du programme, et un temps de traitement relativement court au moment d'appeler "Press" puisque la recherche est plus précise. De plus, comme nous effectuons le traitement dans "Press" nous ne devons construire qu'un petit arbre contenant les différentes prédictions possibles, nous économisons donc beaucoup de mémoire.

Nous avons également implémenté deux boutons + et - permettant de modifier la taille du N-gramme et donc la taille de l'entrée considérée lors de la recherche dans les fichiers.

Extension : Ajouter des bases de données custom **

Pour cette extension, nous avons essayé d'ouvrir une fenêtre de sélection de dossier mais durant nos recherches nous n'avons rien trouvé qui permettait de faire cela en Oz, seul un moyen de sélectionner un fichier particulier était possible. Nous avons donc opté pour le choix d'une fenêtre demandant à l'utilisateur d'entrer le chemin relatif vers le dossier de base de données.

Si le chemin est incorrect, une fenêtre d'erreur s'affiche pour le signaler à l'utilisateur.

De plus, nous avons ajouté la sauvegarde des différents dossiers de base de données renseignés et permet à l'utilisateur de sélectionner le dossier qu'il désire en cliquant sur celui-ci dans la Listbox.

Proposer plus d'un N-gramme *

Cette extension ne requièrait de notre part que l'ajout d'une case permettant de choisir le nombre de mots à proposer. Le programme renvoie donc les X meilleures prédictions. Deux cases + et - permettent de changer la valeur de X .