

CONCLUSION GENERALE

Nous avons présenté dans ce mémoire, une généralisation sur les approches d'analyse intelligente de documents qui utilisent les techniques d'apprentissage automatique, dont le but est la catégorisation thématique des textes multilingues.

Dans le cadre des méthodes d'apprentissage statistique, nous avons présenté et discuter les différentes algorithmes de classification en basant d'une façon plus détaillée sur l'algorithme de l'arbre de décision qui est la méthode qui nous avons utilisé.

La démarche générale est constituée de deux étapes, à savoir :

Le prétraitement du texte : Dans cette phase, la représentation de texte dans un format adapté aux algorithmes d'apprentissage ; on utilise souvent la représentation vectorielle.

Choix d'une méthode d'apprentissage : Nous avons proposé l'algorithme de l'arbre de décision, qui est la méthode la plus populaire et la plus simple.

Un arbre de décisions est une méthode graphique pour analyser des décisions avec risque, des modèles dans lesquels les probabilités associées aux différents états de la nature sont spécifiées. Pour être plus précis, les arbres de décisions ont été conçus pour les problèmes faisant intervenir une séquence de décisions et événements successifs. L'arbre de décision est habituellement représenté comme des décisions ou des événements successifs représentés chronologiquement de gauche à droite.

Comme ce travail était de grande importance, on à rencontré plusieurs difficultés telles que : le problème d'absence des travaux publiés sur les sujets qui concerne la catégorisation des documents multilingues et la classification par arbres de décision et le manque d'ouvrage spécialisées dans le domaine de programmation JAVA(plus précisément interface graphique en java) ,s'ajoute à cela,le non suffisance du temps, la difficulté d'obtenir de l'information relative au domaine dans le moment opportun et en plus le domaine de notre sujet de recherche est très vaste.

Malgré tout cela, nous avons osé nous aventurer dans ce domaine et on peut dire que, vu les résultats obtenus, nous pensons qu'on a quand même pu relever ce défi et par la même occasion apprendre beaucoup de nouvelles connaissances tout au long de la réalisation de ce travail telles que la programmation objet (JAVA), des concepts telles que la classification automatique des documents et les techniques d'extraction des textes, le traitement automatique du langage naturel (en particulier, les langues arabe, anglais et français).

Il serait maintenant intéressant de poursuivre cette recherche, qui a permis de découvrir les avantages de cet algorithme dans le cadre de la catégorisation de texte et en peut proposer comme perspectives, d'élargir notre domaine vers d'autres langues avec plus de corpus de textes, d'intégrer d'autres techniques et

méthodes de classification supervisée et continue avec notre sujet de classification vers l'utilisation de notre méthode d'apprentissage arbres de décision dans la recherche d'information(une petite moteur de recherche), ainsi que toute autre idée jugée utile, réalisable et bénéfique.