Vous créerez un repo Git sur lequel vous placerez votre projet, vous n'aurez plus qu'à me soumettre le lien du git.

L'optimisation pour les moteurs de recherche, aussi connue sous le sigle SEO, inclut l'ensemble des techniques qui visent à améliorer le positionnement d'une page, d'un site ou d'une application web dans la page de résultats d'un moteur de recherche

Certains critères comptent pour le bon référencement d’un site internet. Nous allons en retenir quelqu’un à savoir :

* Lesmots clefs présents pour chaque page du site
* Laprésence de balises alt pour les images
* Lenombre de liens entrants et sortants

Vous allez réaliser un auditeur de référencement SEO pour sites internet à l’aide de fonctions simples

**Etape 1 :** Créer une fonction qui prend en paramètre un texte et qui retourne une liste/dictionnaires de l’ensemble des mots présents dans le texte avec leur occurrence, cette liste sera classée par nombres d’occurrences.

**Etape 2 :** Créer une fonction qui prend en paramètre la structure de données précédentes, ainsi qu’une liste de mots parasites et qui retourne la structure de données privées des mots de la liste parasite.

Le but étant d’enlever les mots parasites comme le, la, les…

**Etape 3 :** Créer une fonction qui récupère les mots parasite dans un fichier parasite.csv et retourne une liste de ses mots.

**Etape 4 :** Testez votre code. Vous devrez pouvoir soumettre une chaine de caractères assez longue. Appelez la fonction de l’étape 1, puis utilisez la fonction de l’étape 3 puis appelez la fonction de l’étape 2 pour obtenir les mots clefs les plus importants de cette page.

**Etape 5 :** Créez une fonction qui prend en paramètre une chaine de caractère au format html et renvoie le même texte mais sans les balises html, vous pouvez utiliser BeautifulSoup.

**Etape 6 :** Ecrire une fonction qui prend en paramètre une chaine de caractère, le nom d’une balise, le nom d’un attribut et retourne la liste des valeurs associées aux balises.

**Etape 7 :** Testez cette fonction pour récupérer dans un code html toutes les valeurs des attributs alt des balises img. Testez cette fonction pour récupérer tous les href des balises a.

**Etape 8 :** Créer une fonction qui prend en paramètre une url et qui extrait le nom de domaine.

**Etape 9 :** Créer une fonction prenant en paramètre une chaine de caractère représentant un nom de domaine, et une liste de valeurs qui sont des url et qui retourne deux listes avec les url qui font partie du domaine et ceux qui n’en font pas partie.

**Etape 10 :**Créer une fonction qui ouvre une page html depuis une url et récupère le texte HTML qui la compose.

**Etape 11 :**Ecrire le programme qui fait l’audit de votre première page

Elle demandera l’url de la page à analyser puis elle affichera les mots clefs avec les 3 premières valeurs d’occurrences, le nombre de liens entrants et le nombre de liens sortant et la présence de balises alt.

Partie 2 :

Vous devrez réorganiser les fonctions précédentes avec des classes et réorganiser votre projet de façon à voir une ihm en tkinter.

L'ihm de votre projet sera la suivante :

une premiere interface où l'utilisateur devra saisir l'url de la premiere page de son site, ainsi que les mots clefs sur lesquels il souhaiterait se faire référencer, ainsi qu'un bouton pour lancer l'analyse et passer à la deuxieme interface.

la deuxieme interface présentera la liste des urls trouvée sur la premiere page du site et pour chacune, on aura les informations suivantes :

 - le nombre de liens sortants présents à cette url (le nombre de liens qui n'appartiennent pas à ce meme site)

- le nombre de liens internes présents à cette url (le nombre de liens qui appartiennent pas à ce site)

- le % de balises alt présents sur le nombre d'images présents

- les 3 premières valeurs des mots clefs pertinents

- si oui ou non, les mots clefs voulus par l'utilisateur sont parmi les 3 premiers.

Vous devrez grâce à des menus :

- pouvoir sauvegarder ce rapport de référencement (dans le format lisible de votre choix)

- pouvoir mettre à jour la liste des mots clefs parasites

bien évidemment, vous devrez pusher sur git votre avancée