

2.4

Prolongements

NSI TLE - JB DUTHOIT

. Typiquement, on peut chercher un mot ou une phrase dans tout le texte d'un roman. Le site <http://www.gutenberg.org/browse/languages/fr> propose les grands classiques de la littérature qui sont tombés dans le domaine public.

On peut par exemple y trouver le texte intégral du roman *Le rouge et le noir* de Stendhal dans l'encodage UTF-8 : <http://www.gutenberg.org/ebooks/798.txt.utf-8>

On pourra alors charger en mémoire ce roman par ces quelques lignes pour chercher ensuite si le motif 'Julien trembla' apparaît quelque part dans le roman :

```
fichier = open('LeRougeEtLeNoir.txt', 'r')
stendhal = fichier.read()
fichier.close()
```

on trouve :

```
>>> len(stendhal)
1055057
```

Remarque

- Pour parcourir le texte de Stendhal, les caractères ASCII ne suffisent pas, il faut faire un pré-traitement plus large!

Combien de fois Stendhal utilise le mot "Julien" dans son livre ?