

# Laboratoire 18 : établissement d'un dictionnaire

- **But**

- Manipulation sur les `string` et les `vector`

- **Travail à réaliser**

- Extraire chaque mot d'un texte français au format UTF8 et le mettre dans un tableau de chaînes de caractères
  - Indications : On supposera que le fichier ne contient que quelques symboles codés en UTF8 sur plusieurs octets : les lettres spéciales du français sont codées sur 2 octets, le premier commençant par `\xC3` ou `\xC5` ; le fichier peut contenir d'autres caractères typographiques, codés sur 2 octets dont le premier est `\xC2`, ou sur 3 octets, dont le premier est `\xE2`
- Afficher les mots dans l'ordre lexicographique (opérateur `<` standard), en précisant pour chaque mot le nombre de ses occurrences dans le texte.
- Quel est le personnage le plus cité dans *La Comédie Humaine* ?
- Quel est le mot le plus long ?

- **Délai**

- 2 séances