Consignes du projet - 1

- Écrire un premier script Python permettant d'aspirer (collecter) les entités médicales de type **noms de médicaments par substance active** de A à Z, à partir des 26 pages HTML du dossier « VIDAL » que je vous ai mis en pièce-jointe.
- Générer en sortie un dictionnaire au format **.dic** (format DELAF vu en cours 3) encodé en **UTF-16 LE avec BOM** (UCS-2 LE BOM).
- Ce dictionnaire **doit s'appeler** « **subst.dic** » et doit contenir les médicaments par substance active des 26 pages HTML du dossier « VIDAL ».
- Chaque entrée lexicale de ce dictionnaire doit être suivie par les informations (codes) **,.N+subst**
- L'information **N** est de type grammatical et l'information **subst** est de type sémantique.
- Vous devez donc obtenir une sortie ayant le format DELAF-UNITEX suivant :
    - **abacavir,.N+subst**
    - **abatacept,.N+subst**
    - **abciximab,.N+subst**
    - **abiratérone,.N+subst**
    - **…….**
- Pour faire votre aspiration en local sur votre machine, vous devrez installer une plate-forme de développement Web, comme par exemple : **WampServer, XAMPP** ou **EasyPHP-DevServer**, qui contiennent, entre autres, un serveur Web Apache.

Remarque : l'encodage des pages HTML du dossier « VIDAL » ne doit pas être modifié.