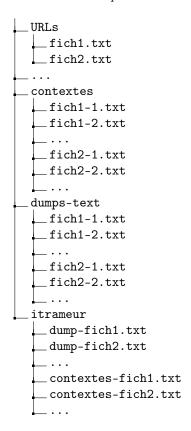
# Générer une base iTrameur

#### Mot d'introduction

Le but de cette feuille est de créer **un nouveau script** afin de formater votre corpus afin de le rendre lisible par iTrameur. iTrameur <sup>1</sup> est un outil d'exploration de corpus et de textométrie.

Il faudra créer divers fichiers au cours de ce TD, à ranger selon l'architecture de dossiers suivante, qui enrichit celle déjà existante évoquée dans les feuilles précédentes. Les noms de type fich1-1.txt reprennent le nom du fichier d'URL correspondant :



Note 1 Nous travaillerons ici sur les fichiers de dump et de contextes. Il ne sera pas demandé d'avoir une connexion internet.

## Exercice 1 Exemples de fichier base iTrameur

Un fichier d'exemple pour iTrameur (appelé base iTrameur) est disponible sur le git. Vous pourrez vous en servir comme base pour créer la structure de votre fichier. Un exemple proche du minimum que vous pouvez copier-coller est le suivant. Vous pourrez noter la présence de symboles "§" après les balises page, ces symboles sont importants car ils sont les marqueurs de section par défaut dans iTrameur, pensez à les mettre.

```
<lang="fr">
<page="fr-1">
<text>|ci, le contenu du fichier fr-1.</text>
</page> $
<page="fr-2">
<text>|ci, le contenu du fichier fr-2.</text>
</page> $
</lang>
</lang>
<lang="en">
<page="en-1">
<text>|ci, le contenu du fichier en-1.</text>
</page> $
</lang>
</page> $
</lang>
</page> $
</lang>
```

<sup>1.</sup> http://www.tal.univ-paris3.fr/trameur/iTrameur/

## Exercice 2 Travail sur les dumps textuels

Créez avant de lancer les scripts suivants le dossier itrameur.

Créez un script make\_itrameur\_corpus.sh. Ce script prendra en argument :

- un dossier
- un nom de base qui correspond au nom du fichier URL sans son extension (fich1 ou fich2 dans l'exemple plus haut)

Le script devra créer une base iTrameur qui prendra la forme indiquée dans l'Exercice 1. Pour ce faire, il y aura deux étapes :

- 1. créer un fichier par langue
- 2. concaténer les différents fichiers pour obtenir une base globale pour toutes les langues

Pour 1., il faut itérer sur les différents fichiers dump (1 par URL), et les intégrer à une balise page et mettre le contenu textuel (dump) dans une balise text, comme indiqué plus haut. Pour le nom de base fich1, le fichier écrit aura alors le nom dump-fich1.txt et sera placé dans le dossier itrameur.

Pour 2., il suffit de concaténer l'ensemble des fichiers dans un fichier final appelé dump.txt

### Exercice 3 Travail sur les contextes

Refaites les manipulations précédentes, mais en les appliquant aux fichiers de contexte. Ajoutez le tout dans le script make\_itrameur\_corpus.sh. Les fichiers créés devront avoir des noms de type contexte-fich1.txt et le fichier final sera contexte.txt dans le dossier itrameur.

#### Exercice 4 Travail sur iTrameur

En utilisant la feuille de travail sur iTrameur, chargez votre base sur le iTrameur et analysez votre corpus. Une fiche de travail et un turotiel sont disponibles sur icampus dans la section Ressources pour le cours, rubrique Documentation iTrameur.