

LECTURE DE PARAMÈTRES

TRAITEMENT FICHER

for fichier in \$(ls \$DOSSIER_URLS)

AFFICHAGE TABLEAU

TRAITEMENT URL

for LINE in \$(cat \$DOSSIER_URLS/\$fichier)

RÉCUPÉRATION PAGE ASPIRÉE

Commande CURL < charset= « x »>

OPTION A



UTF-8

TRAITEMENTS:

- Téléchargement URL avec lynx
- Motif
- Extraction contexte
- Fréquence motif
- Index
- Dump total

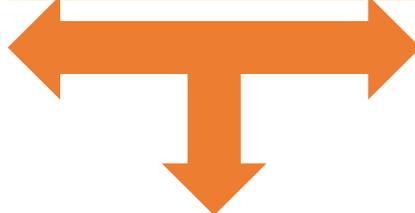
OPTION A

UTF-8

OPTION B

!=UTF

DÉTECTION ENCODAGE



PAS D'ENCODAGE

PAGE ASPIRÉE EXISTE?

OUI ✓

file -i?

OUI

NON



OPTION B

Encodage != UTF-8

VÉRIFIER LISTE ICONV

OUI ✓

NON X

ICONV

TRAITEMENTS

