# TransliTAL

Corriger et (rétro)-translittérer les notices catalographiques de documents en écritures non-latines du Sudoc à l'aide de modèles de langues de traitement automatique des langues (TAL)

[ א ט a ]

## *Objectifs du projet*

- **résoudre les ambiguïtés d'identification** des langues des notices
- repérer et **corriger les erreurs de transcription ou de translittération** des champs en bi-écriture
- assurer la présence de l'écriture originale pour les notices en translittération seule
- proposer un **outil de translittération ou rétro-translittération automatique** pour les catalogueurs du réseau

## *Terrains d'expérimentation*

Mise en œuvre **en partenariat avec le laboratoire ERTIM** de techniques d'**apprentissage automatisé** basé sur des modèles de langues pour des corpus de notices :

- en langues à **écriture arabo-persane** : arabe, persan, ourdou, turc…
- en langues à **écriture cyrillique** : russe, ukrainien, biélorusse, serbe, bulgare et autres langues cyrillisées
- en langues utilisant les **sinogrammes** : chinois, japonais, coréen, viêtnamien chữ nôm

[ あ A ]

[ 阿 ]

## *Les 4 temps du projet*

1/ préparation des corpus
2/ entraînement des modèles de langues en **apprentissage non supervisé et semi-supervisé**
3/ développement d'une **interface utilisateur**
4/ mise en production et première vague de **modifications en masse** sur les notices du Sudoc

abes ; agence bibliographique de l'enseignement supérieur

BULAC [গ্রন্থাগার] [جامعية] [ ЈЕЗИЦИ ] [+ᴏ:ᴏᴄᴏ]

inalco ERTIM Equipe de recherche textes, informatique, multilinguisme