



TRANSKRIBUS

Atelier ADEMEC

5 mars 2019

Alix Chagué

Transkribus c'est quoi ?

Une solution développée

- depuis 2013 grâce à des financements européens (H2020).
- par plusieurs équipes rassemblées par le projet [READ](#) (*Recognition and Enrichment of Archival Documents*).
- en partie seulement en *open-source*

--

OCR : Optical Character Recognition

HTR : Handwritten Text Recognition

OCR et HTR = transcription automatique

Transkribus c'est quoi ?

1. Une plateforme qui permet de

- charger des images
- transcrire manuellement
- transcrire automatiquement
- entraîner un modèle de transcription

2. Une interface en ligne

- transkribus.eu/r/read/login/

3. Une API REST

- transkribus.eu/wiki/index.php/REST_Interface

Qu'attendre de Transkribus ?

Au pire	Au mieux
une plate-forme de transcription manuelle	un outil pour automatiser la transcription
	

0% d'erreur de reconnaissance (CER) ça n'existe pas

automatiser ne sert à rien sur un corpus petit et/ou hétérogène

Puis-je entraîner un modèle ?

- compter transcrire manuellement au moins 75 pages d'une même main (voire plus pour des textes difficiles)
- compter gagner du temps à la transcription mais en passer plus au pré-traitement des images (*recadrage, contraste, ...*)
- compter gagner du temps à la transcription mais en passer plus à la correction post-OCR

--

Si je juge que le pré/post-traitement ne me fait pas passer plus de temps à obtenir le même résultat qu'en transcrivant à la main, alors oui !

Liens utiles :

Transkribus Wiki :

- https://transkribus.eu/wiki/index.php/Main_Page

Transkribus FAQ :

- https://transkribus.eu/wiki/index.php/Questions_and_Answers

Tutoriel de base :

- https://transkribus.eu/wiki/images/7/77/How_to_use_TRANSKRIBUS_-_10_steps.pdf