

# Extraction et catégorisation automatique de contenu de manuels scolaires

## Contexte

Le projet ANR MALIN a pour objectif de rendre utilisables les manuels scolaires numériques par les enfants en situation de handicap dont l'inclusion dans les établissements scolaires ordinaires a été posée par la loi du 11 février 2005. En France, le manuel scolaire est un support pédagogique emblématique quasi systématiquement utilisé en classe afin d'accompagner de manière progressive et structurée l'acquisition des connaissances et compétences définies dans les programmes scolaires officiels de l'Éducation Nationale. Les manuels numériques actuellement disponibles nécessitent d'être adaptés pour être utilisés par les enfants en situation de handicap. Ces adaptations concernent aussi bien les aspects techniques que pédagogiques. Dans la plupart des cas, des parties seulement des manuels sont adaptées et les délais de livraison peuvent être de plusieurs mois. Ces contraintes ne permettent pas de rendre efficiente l'inclusion scolaire des enfants en situation de handicap. L'objectif du projet ANR MALIN est donc de développer des solutions techniques afin d'aboutir à **l'automatisation de l'adaptation des manuels scolaires numériques pour les rendre accessibles** (accès, traitement et interaction avec les contenus) **aux élèves en situation de handicap**.

Dans ce contexte, le/la stagiaire travaillera sur l'extraction des consignes des exercices ainsi que sur leur identification / catégorisation. L'objectif du stage consiste notamment à identifier les consignes de type "choix proposés" dans les différents exercices. Ces choix proposés pouvant prendre plusieurs formes (voir exemples en figure 1), le/la stagiaire développera, à partir d'une chaîne d'extraction déjà existante et de données annotées manuellement, des techniques génériques, fondées sur des approches de traitement automatique des langues et tirant partie de l'efficacité des technologies d'apprentissage profond.

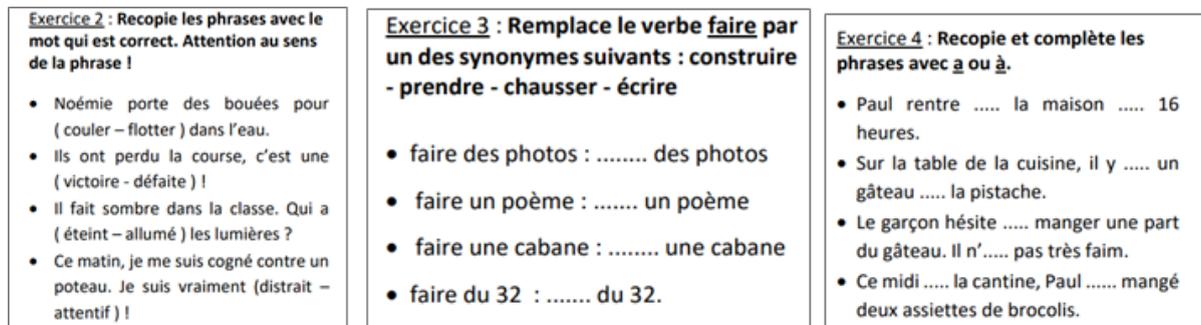


Figure 1

Ce stage s'inscrit dans le cadre d'un projet ANR et pourra se poursuivre par un doctorat de 3 ans. Le projet ANR repose sur une collaboration entre quatre laboratoires : LISN (Université Paris Saclay), MISC (Ecole CentraleSupélec), CEDRIC (CNAM), Inserm 1284 (CRI, Université de Paris). Le stagiaire travaillera en interaction avec deux autres stagiaires associés au projet.

## Compétences

- master et/ou ingénieur en informatique avec une spécialisation dans au moins un des domaines suivant:
  - traitement automatique des langues
  - apprentissage automatique
- maîtrise de Python (langage de prédilection du projet)
- maîtrise de l'anglais (écrit et oral)

La connaissance de bibliothèques d'apprentissage sera appréciée.

## Informations générales

Lieu de travail : Laboratoire Interdisciplinaire des Sciences du Numérique (LISN) à Orsay

Durée du contrat : 6 mois

Date d'embauche souhaitée : entre janvier et mars 2022

Quotité de travail : Temps complet

Gratification de stage d'organisme public

Niveau d'études souhaité : Bac+5

**Contact** : Pour postuler, merci d'envoyer un CV et une lettre de motivation à Camille Guinaudeau ([guinaudeau@limsi.fr](mailto:guinaudeau@limsi.fr)), Olivier Pons ([olivier.pons@lecnam.net](mailto:olivier.pons@lecnam.net)) et Caroline Huron ([caroline.huron@paris-cri.org](mailto:caroline.huron@paris-cri.org)).