Extraction et enrichissement de contenu multimodal
Application au cas des manuels scolaires

Contexte de la thèse

Le projet ANR MALIN a pour objectif de rendre utilisables les manuels scolaires numériques par les enfants en situation de handicap. En effet, les manuels numériques actuellement disponibles nécessitent d'être adaptés pour être accessibles à ces enfants. Ces adaptations concernent aussi bien les aspects techniques que pédagogiques. Dans la plupart des cas, les manuels sont adaptés de façon artisanale 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**.

Le projet ANR repose sur une collaboration entre quatre laboratoires : LISN (Université Paris Saclay), MICS (Ecole CentraleSupelec), CEDRIC (CNAM), Inserm 1284 (CRI, Université de Paris). Le doctorant ou la doctorante travaillera en interaction avec des stagiaires de master, des ingénieurs et un autre doctorant associés au projet.

Sujet de la thèse

Le premier objectif est de concevoir des approches d'extraction automatique de la structure d'un manuel (leçons, blocs d'exercices [eux-mêmes composés de consignes, enoncés, exemples…], memo, synthèse...) et de son contenu multimédia (textes, images, dessins, graphiques, équations, courbes...) à partir des fichiers fournis par les éditeurs (ceux-ci sont le plus souvent au format pdf). Plusieurs approches seront à envisager : une approche d'adaptation et d'enrichissement de systèmes de structuration automatique de documents textuels (segmentation thématique, segmentation discursive) prenant en compte la spécificité et la multi-modalité des données traitées et une approche basée sur le traitement automatique des images visant à identifier les différents blocs en se basant sur les caractéristiques de l'image, connue sous le nom de « Document Layout Segmentation and Analysis » [1, 2]. Des approches récentes d'apprentissage profond seront testées sur des jeux de données annotées manuellement afin d'adapter des modèles existants et obtenir des résultats d'extraction satisfaisants.

Le second objectif est d'analyser le contenu de chaque bloc extrait dans l'étape précédente afin de les catégoriser en activités pédagogiques. Ainsi pour chaque exercice, il faudra déterminer quelle(s) activité(s) pédagogique(s) devra(ont) être mise(s) en œuvre pour le réaliser. Dans cet objectif, le doctorant ou la doctorante devra développer des techniques d'apprentissage spécifiques novatrices, supervisées ou non, à la rencontre entre le traitement du langage naturel et la linguistique d'une part et l'analyse de données multimédia d'autre part [3, 4]. Dans ce cas, les modalités visuelles et textuelles seront représentées dans un espace commun pour effectuer une classification multimodale. L'une des pistes à explorer consiste à employer des modèles appris sur des données et des problématiques similaires en français et de travailler sur l'adaptation (fine-tuning) de ces modèles à partir d'un petit jeu de données annotées [5, 6, 7].

Compétences

- Master en informatique ou TAL avec une spécialisation dans au moins un des domaines suivants :
    - traitement automatique des langues
    - apprentissage automatique

● Maîtrise de Python (langage de prédilection du projet)
La connaissance des principales librairies d'apprentissage sera appréciée.

Informations générales

Lieu de travail : Laboratoire CEDRIC du CNAM à Paris et Laboratoire Interdisciplinaire des Sciences du Numérique (LISN) à Orsay
Durée du contrat : 36 mois
Début de la thèse : octobre/novembre 2022
Contact : Pour postuler, merci d'envoyer un CV, les notes de M1 et M2 et une lettre de motivation à Camille Guinaudeau (guinaudeau@limsi.fr), Olivier Pons (olivier.pons@lecnam.net) et Caroline Huron (caroline.huron@cri-paris.org).

Références

[1] XU, Yiheng, LI, Minghao, CUI, Lei, *et al.* Layoutlm: Pre-training of text and layout for document image understanding. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020.
[2] Bakkali, S., Ming, Z., Coustaty, M., & Rusiñol, M. (2020). Visual and textual deep feature fusion for document image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.
[3] LIANG, Tao, LIN, Guosheng, WAN, Mingyang, *et al.* Expanding Large Pre-Trained Unimodal Models With Multimodal Information Injection for Image-Text Multimodal Classification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
[4] CHOI, Jun-Ho et LEE, Jong-Seok. EmbraceNet: A robust deep learning architecture for multimodal classification. *Information Fusion*, 2019, vol. 51, p. 259-270.
[5] HOWARD, Jeremy et RUDER, Sebastian. Universal Language Model Fine-tuning for Text Classification. In: *ACL 2018-56th Annual Meeting of the Association for Computational Linguistics*. 2018.
[6] MARTIN, Louis, MULLER, Benjamin, SUÁREZ, Pedro Javier Ortiz, *et al.* CamemBERT: a Tasty French Language Model. In: *ACL 2020-58th Annual Meeting of the Association for Computational Linguistics*. 2020.
[7] H. Le, L. Vial, J. Frej, V. Segonne, M. Coavoux, B. Lecouteux, A. Allauzen, B. Crabbé, L. Besacier, D. Schwab. FlauBERT: unsupervised language model pre-training for French. In: *Proceedings of the 12th Language Resources and Evaluation Conference (2020)*, pp. 2479-2490