

## Hekzam : compromis précision/performance de l'OCR

### Contexte

Hekzam est un ensemble de logiciels en cours de conception qui a pour objectif de permettre l'écriture d'examens papiers qui peuvent être corrigés automatiquement ou pseudo-automatiquement. Voici la manière typique dont Hekzam est utilisé pour faire et corriger un examen.

1. L'équipe enseignante crée un formulaire de sujet d'examen grâce aux logiciels fournis de Hekzam
2. Les apprenant-e-s passent leur examen en complétant le formulaire
3. L'équipe enseignante scanne les formulaires complétés
4. L'équipe enseignante utilise les logiciels fournis de Hekzam pour obtenir
  - Des notes (et analyses détaillées) pour les questions à correction complètement automatisée
  - Les saisies des apprenant-e-s (images, booléens, nombres...) pour les autres types de questions

Hekzam permet de définir plusieurs types de champs dans un formulaire. Certains de ces champs nécessitent de faire de la reconnaissance optique de caractères (OCR) pour transformer la saisie de l'apprenant-e en donnée utilisable par un traitement automatisé. On va ici s'intéresser à deux champs.

- Champ de type chiffre unique. Par exemple une séquence de 8 champs pour un numéro étudiant.
- Champ de type suite de mots. Par exemple pour indiquer son prénom + nom.

### Objectifs du projet

Le but principal du projet est d'évaluer les compromis entre précision et performance d'algorithmes et de technologies d'OCR utilisables par Hekzam. Il est à noter que Hekzam est conçu dans une démarche de sobriété numérique et a pour but de pouvoir être exécuté sur des ordinateurs très peu puissants.

Une étude bibliographique sera réalisée sur cette question, pour avoir un aperçu de l'état actuel du problème côté publications scientifiques. En parallèle, une étude expérimentale sera réalisée pour évaluer la précision et le coût de traitement de différentes technologies d'OCR libres et open source. Des modèles de machine learning *ad hoc* pourront être entraînés et évalués à condition que l'apprentissage soit très peu coûteux et que ces modèles soient utilisables sur un ordinateur très peu puissant.

### Candidature

Contactez-moi par mail avec les informations suivantes.

- Très courte motivation par rapport au sujet (2 phrases)
- Bulletins de notes (master et licence)
- CV court (parcours d'études, expériences professionnelles, compétences)