

# Projet de développement

## Sprint 4

**Equipe** : Romain Le Talour, Corentin Leconte, Thibaut Quilleré.

Pour ce projet de développement, la méthodologie de travail est basée sur l'agile avec une approche SCRUM.

**Maitre SCRUM** : Romain Le Talour

**Lien du GitHub** : <https://github.com/genielog-ue/pdftotxt.git>

Dans ce quatrième sprint, la répartition des tâches s'est faite comme à l'accoutumée. Tout le monde s'est mis sur le code, en particulier Corentin. Je me suis aussi chargé de modifier le README.md sur GitHub car les consignes ont fait que les fonctionnalités du parseur ont changé.

Nous avons eu quelques difficultés à implémenter le menu et plus particulièrement le choix de conversion en TXT ou XML ou les deux après avoir choisi le ou les PDF à traiter.

Le menu s'affiche bien évidemment dans le terminal et l'utilisateur utilise le prompt pour faire son choix.

Nous avons mis des numéros en face de la liste des PDF présents dans le répertoire choisi, cela facilite l'utilisateur dans son choix. Il n'a pas besoin de taper le nom du PDF. Il suffit simplement de taper les numéros des PDF entre parenthèses, séparé par un virgule ou non ( *(1,2,3)* ou *(1 2 3)* ).

Ceux-ci sont classés dans l'ordre de lecture du répertoire par le programme.

Autrement les fonctionnalités demandées sont présente et fonctionnent bien.

Vous trouverez sur le GitHub (branche sprint4) les captures d'écran des artefacts comme d'habitude ainsi que le fichier main.py et le dossier Corpus\_2021.

.....

Je tenais à revenir sur le compte rendu du sprint 3. Je me suis trompé et ai envoyé le compte rendu du sprint 2. Je joins ci-après le rendu du sprint 3. Merci.

# Projet de développement

## Sprint 3

Equipe : Romain Le Talour, Corentin Leconte, Thibaut Quilleré.

Pour ce projet de développement, la méthodologie de travail est basée sur l'agile avec une approche SCRUM.

Maitre SCRUM : Romain Le Talour

Lien du GitHub : <https://github.com/genielog-ue/pdftotxt.git>

Dans ce troisième sprint nous nous sommes premièrement penchés sur la correction du bug et amélioration du code.

La consigne de ce troisième sprint et donc la tâche la plus importante était de rajouter un sortie XML dans le parser. Pendant l'entretien de la séance du sprint 3, il nous a été souligné que des éléments étaient manquants lors de nos rendus tels que des preuves de nos mêlées etc...

C'est maintenant chose faites et vous pourrez retrouver des captures d'écrans de nos mêlées qui se déroulent environ tous les 2/3 jours. La répartition des tâches était plus ou moins comme au sprint 2. Corentin s'est en grande parti occupé de corriger le sprint 2 et nous avons réfléchi à ajouter des méthodes pour pouvoir bien séparer les fonctionnalités et dans le but d'optimiser le programme.

Nous avons aussi implémenté le Regex dans notre code pour nous faciliter le travail de séparation et de recherche de paragraphes précis dans un PDF (ici ces PDF sont dans leur globalité scientifique, ce qui nous donne un pattern et donc facilite la recherche).

Bien sûr les fonctionnalités sont à chaque fois testées, dans les prochains sprints nous nous pencherons davantage à push des fichiers de test unitaires.