

Sprint 3 - Parseur d'articles scientifiques en format texte

Repo Github : <https://github.com/elliottlepine/scrum-tp/tree/GL-1-Veille-techno-pdf2text>

Pour un confort optimal, visionner ce document à cette adresse : <https://github.com/elliottlepine/scrum-tp/blob/version-2/README.MD>

Elliott Lépine, Lyes Douki, Yacine Trousselle

Lancer l'application

```
python3 ./app/index.py --text --input [chemin vers PDF] --output [chemin vers TXT]  
# OR  
python3 ./app/index.py --xml --input [chemin vers PDF] --output [chemin vers XML]
```

Options

- t / --text : Convertir au format TXT
- x / --xml : Convertir au format XML
- i / --input : Chemin vers le fichier d'entrée
- o / --output : Chemin vers le fichier de sortie
- m / -menu : Sélectionner les fichiers via un menu

Compte rendu

Analyse du besoin (17/11/2022 - 9h)

Voir <https://github.com/elliottlepine/scrum-tp/blob/master/comprehension.md>

Planification Sprint 1 (17/11/2022)

Réunion de planification avec les 3 membres de l'équipe.

Remplissage du backlog avec des tickets atomiques respectant le cahier des charges.

Paramétrage du premier sprint.

Tickets GL-1 et GL-2 à faire.

Voir <https://github.com/elliottlepine/scrum-tp/blob/master/backlog.md>

Rétrospective du sprint 1 et planification du sprint 2 (05/12/2022)

Les tickets du sprint 1 n'étaient pas tout à fait terminés en raison des contraintes de temps.

Ils passent sur le sprint 2, et on ajoute les tickets GL-4 et le GL-5. Le ticket GL-8 est quant à lui annulé.

L'équipe a décidé d'un commun accord de développer le logiciel en Python et d'utiliser l'utilitaire pdftotext, suite à la veille technologique réalisée lors du sprint 1.

Simili daily-scrum de mi-séance (05/12/2022)

L'équipe s'est réunie au bout d'1h30 pour faire le point sur l'avancement des tâches respectives.

Le choix de l'utilitaire s'est finalement porté sur pdftotext, puisque les essais réalisés avec pdf2text n'étaient pas lisibles lorsque le document contenait 2 colonnes.

Il a été décidé de passer toute l'équipe sur une première implémentation de pdftotext dans l'app Python.

Un Trello a été créé pour pouvoir gérer le backlog de manière plus simple

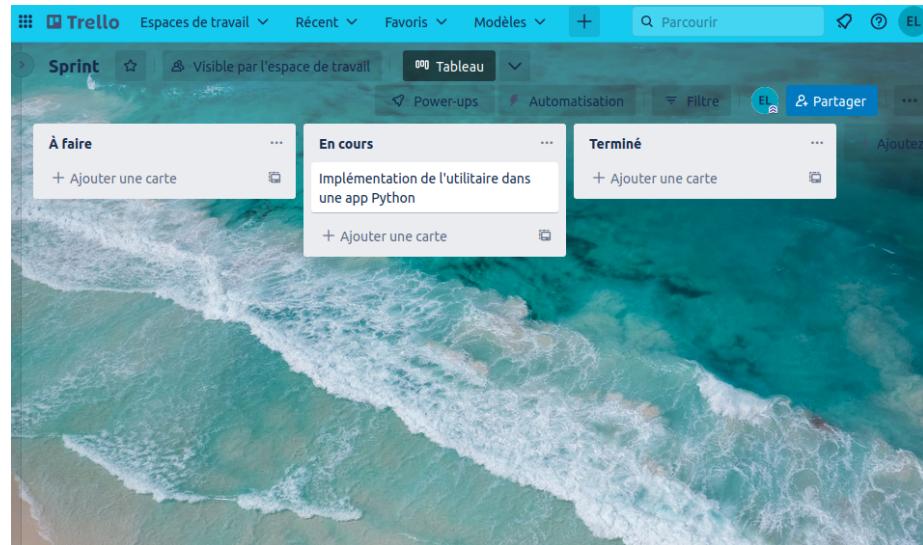


Figure 1: Trello

Rétrospective du sprint 2 et planification du sprint 3

Les tâches d'extraction n'ont pas pu être effectuées lors du sprint 2.

Elles ont été transférées sur le sprint 3, en plus de la tâche de conversion en XML.

Ces changements ont été reportés sur le Trello du projet.

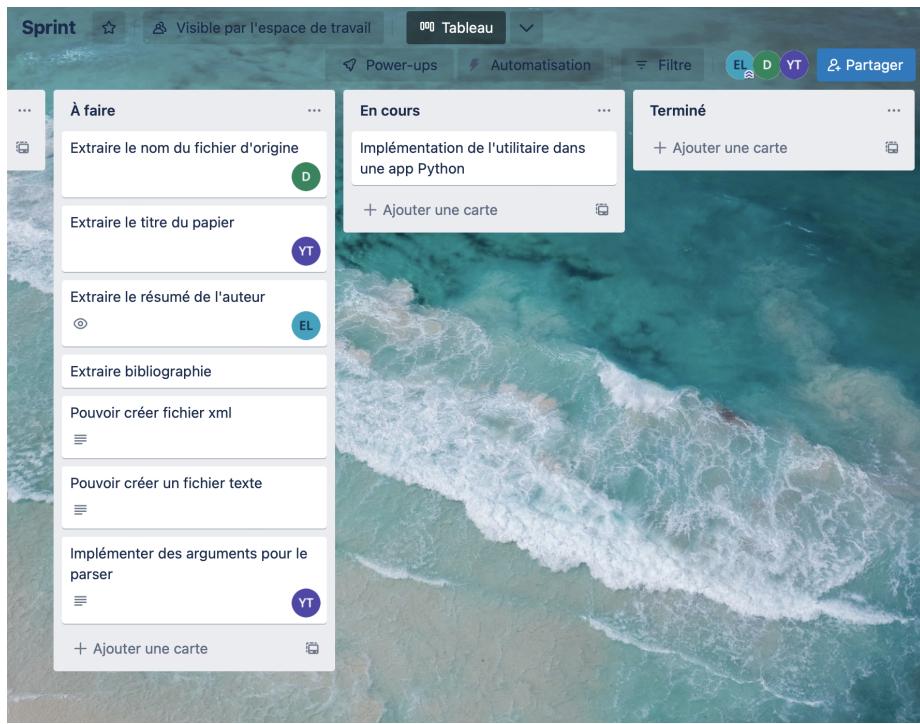


Figure 2: Trello sprint 3

Rétrospective du sprint 3 et planification du sprint 4 (14/12/2022)

Il restait 2 tickets du sprint 3 : la création de fichiers XML, et l'extraction de la bibliographie. Ils ont tous les 2 été passés sur le sprint 4.

Des tickets concernant l'extraction de nouvelles informations et la création d'un menu ont été ajoutés au sprint 4 et répartis entre les différents membres de l'équipe.

Les changements ont été reportés sur le Trello du projet.

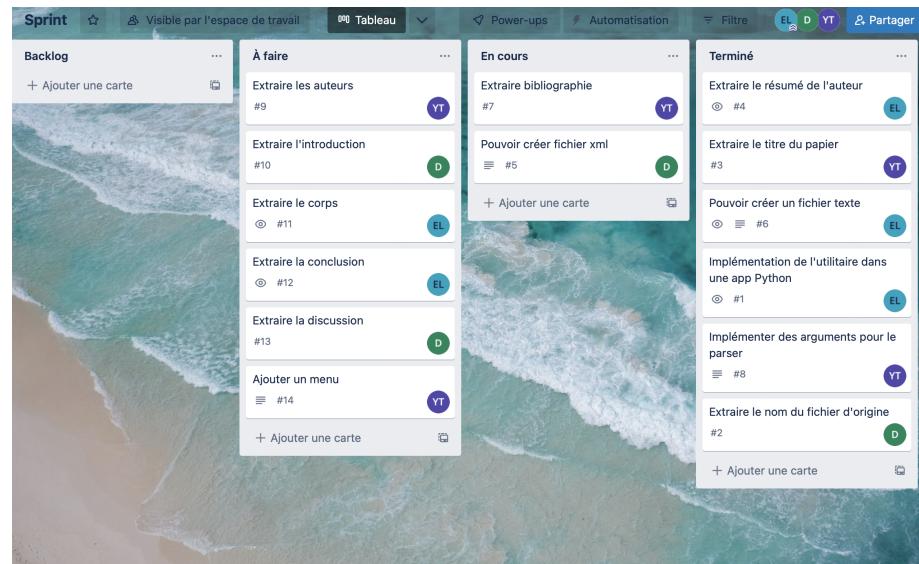


Figure 3: Trello sprint 4