

Rapport Sprint 2: Génie logiciel – Scrum

Temps donné total (3h00).

Objectif : Récupérer le nom des fichiers, le titre et le résumé de chaque fichier pdf dans des fichiers texte.

Product backlog :

À faire	Priorité	Estimation
Déterminez le meilleur langage à utiliser	Basse	30 min
Convertir les fichier pdf d'un dossier en fichier Texte et les placer dans un sous dossier	Moyen	1h
Récupérer des information dans les fichiers textes convertie et les mettre dans de nouveaux fichiers textes	Haute	1h30

1. Planification du Sprint:

L'équipe s'est réunie pour discuter des exigences et de l'objectif du sprint. Nous avons choisie de changer la méthode de notre sprint 1 afin d'automatiser la conversion des pdf en fichier texte et afin de tout généraliser sur python pour plus de simplicité. On a donc décidé de prendre le langage Python pour réaliser nos fonctions car c'est le langage que l'on maîtrise le mieux. Dylan va s'occuper de la fonction Python qui récupère les noms des fichiers pdf d'un dossier, les lit et écris le contenu en texte brute dans des fichiers textes du même noms des fichiers pdf, en les plaçant dans un sous dossier crée. Enzo va s'occuper de la fonction Python qui prend en argument le résultat de la fonction de Dylan, puis récupère le nom, le titre et le résumé de chaque fichier, et renvoie une liste de fichiers textes contenant toutes ces informations. Et enfin Romain va s'occuper du rapport.

2. Mise en œuvre du Sprint:

Scrum 1: réunion pour définir les objectifs et répartir les taches.

1^{er} étape : Choix du langage.

2^e étape : Code de conversion des fichiers pdf en txt

Scrum 2: Au bout d'1h30, réunion pour discuter de l'avancement, des obstacles et, si nécessaire, ajuster les plans.

3^e étape : fonction de récupération des données des fichiers texte.

4^e étape : Évaluation des résultats.

Scrum 3: Au bout de 2h30, réunion pour discuter de se qu'il nous manque ou non et préparer le rendu.

5^e étape : Rendu du Sprint 2

3. Revue de Sprint:

Évaluation des Résultats:

- Les 10 fichiers PDF ont été convertis en plain texte dans un sous dossier grâce à la fonction python « convert.py ».
- Extraction du nom des fichiers, du titre et du résumer des 10 fichiers plain texte

Démonstration: Présentation des fichiers sortie après extraction.

Rétrospective: Réflexion sur le sprint, recherche des points forts et améliorations potentielles pour les prochains sprints.

4. Conclusion du Sprint:

Résultats:

- La conversion des pdf en fichier texte brute se passe sans encombre (voir photo_1).
- L'analyse et la récupération des données des fichiers textes se passe bien dans la majorité mise à part le fichier mikheev et compression. Sur le fichier mikheev on a du récupérer en plus tout le début afin de pouvoir avoir le résumer. Sur le fichier compression la dernière phrase est coupé après « lan- »(voir photo_2).

Les connaissances acquises :

- utilisation Bibliothèque Python PyPDF2 qui convertit des fichiers pdf en txt
- parser un fichier txt en python