

# **Rapport de milieu de Sprint**

## **Sprint numéro : 2**

### **Tâches finies :**

- Analyse du fichier et en particulier de son en-tête
- Extraction du nom du fichier d'origine
- Extraction du titre du papier
- Extraction du résumé (abstract)
- Tests sur les différents fichiers « .pdf » du corpus
- Création de tests (unitaires notamment)

### **Difficultés rencontrées :**

- Problème d'extraction du titre sur quelques fichiers « .pdf » notamment à cause des titres sur plusieurs lignes
- Problème de caractères non-lisibles
- Des auteurs peuvent être sur une seule et même ligne
- Mauvaise extraction de l'abstract/résumé pour certains « .pdf »

### **Tâches à venir :**

- Extraction du/des nom(s) du/des auteur(s)
- Améliorer l'extraction de l'abstract et des auteurs
- Améliorer les tests unitaires

**Ressenti de l'équipe :** Aucune fatigue a été ressentie durant la première semaine de Sprint. Les membres de l'équipe sont confiants pour respecter les délais.

Certains « .pdf » notamment ceux de M. Kessler sont les plus compliqués à traiter de par leur mise en page particulière.

**Conclusion :** En conclusion de ce rapport de ce milieu de Sprint, l'équipe va respecter les délais et avance à un bon rythme.