



UE: Génie logiciel

# Sprint 2: Génie logiciel

Présenté par : BANFI Débora (Srum master)  
CORNEN Titouan  
FREHOU Soufiane  
IDRISSOU Fataï  
MOUNDIR Narjis



# Résumé Sprint I:

Choix de l'outil de conversion:

PDFTOTEXT

PDF2TEXT

Comme vu lors du Sprint 1 et après avoir défini les avantages et inconvénients de ces deux outils, nous avons décidé de choisir PdfToText qui nous paraît être la meilleure option.





# Sprint 2:

Parser d'un article scientifique  
en format texte







# Le Contexte



- Créer un système qui prend en entrée le nom d'un dossier contenant des fichiers PDF et créer un sous-dossier dans lequel se trouvera les fichiers .txt correspondant aux fichiers PDF.
- Les fichiers .txt devront respecter une forme, ils devront préciser les différentes sections du fichier (titre, auteur, abstract...).

## Les langages utilisés:

- Python
- Bash

# Performance de temps d'exécution

Pour chaque langages choisis

	BOUCLE FOR DE 0 À 1000	OUVERTURE ET LECTURE D'UN FICHIER
Python	Real 0m0.037s user 0m0.014s sys 0m0.015s	Real 0m0.083s user 0m0.029s sys 0m0.024s
Bash	real 0m0,029s user 0m0,006s sys 0m0,008s	real 0m0,022s user 0m0,004s sys 0m0,007s

Ces languages ont été choisi en fonction des performance d'exécution mais aussi en fonction de leur maîtrise par chaque membre du groupe.

des différentes tâches:

- Parseur :**  
C++  
ou  
PYTHON  
↳ déterminer les différentes parties du texte (titre, auteur, abstract...)  
↳ déterminer le contenu du fichier
- Le fichier .txt :**  
BASH  
ou  
PYTHON  
↳ créer le fichier .txt correspondant  
↳ Extraire :
  - les différentes parties
  - le contenu du fichier PDF
- Boucle :**  
BASH  
ou  
PYTHON  
↳ Faire une boucle pour exécuter les scripts sur tous les fichiers d'un dossier

1

**CRÉATION D'UN PARSEUR**

Implementer un parseur afin d'identifier les différentes parties du fichier (titre, abstract, auteur...)

2

**CRÉATION DES SOUS-DOSSIERS**

Implementer un script qui va créer un sous-dossier (si il n'existe pas déjà) qui contiendra les fichiers .txt, dans le cas contraire il le supprimera.

3

**CRÉATION DES FICHIERS .TXT**

Implémenter un script qui nous permettra de créer et d'écrire les fichiers .txt.

# Tâches générales à réaliser



# ➤ Répartition des tâches

Les tâches ont été réparties en fonction des préférences des langages de chacun

EN PYTHON

EN BASH

- Parser : **NARJIS**

- ↳ déterminer les  $\pm$  parties du fichier (titre, auteurs...)
- ↳ écrire dans le fichier.txt

- Récuper le nom du fichier PDF
- Créer un fichier.txt correspondant au fichier PDF (avec le même nom)

- Bouche :

**FATAH**

- ↳ créer un sous-dossier, si il existe pas le supprimer et on le recrée

**DÉBORAH**

- ↳ faire un script qui va exécuter tous les autres scripts sur tous les fichiers du dossier

**TITOUAN**

**SOUFIANE**