



UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE

C E N T R E
D'ENSEIGNEMENT
ET DE RECHERCHE
EN INFORMATIQUE



Licence D'informatique
Systèmes et Réseaux Informatiques, Ingénierie Logicielle
Parseur d'articles scientifiques en format texte

Rendu TP2

3 décembre 2018
Groupe 3, Groupe SCRUM 4

Nizar Rezaigui (SCRUM Master)
Thibaut Chastelliere
Florian Feuillepain
Zinedine Makhlouf

CERI - LIA
339 chemin des Meinajariès
BP 1228
84911 AVIGNON Cedex 9
France

Tél. +33 (0)4 90 84 35 00
Fax +33 (0)4 90 84 35 01
<http://ceri.univ-avignon.fr>

Sommaire

Titre	1
Sommaire	2
1 Introduction	3
2 Choix du langage	3

1 Introduction

L'objectif de cette séance est de choisir quel langage nous allons utiliser pour répondre aux attentes du clients. Nous devons donc le choisir en fonction de la maîtrise des langages par les membres de l'équipe et en fonction de la vitesse d'exécution de la commande *pdftotext* (Lors du précédent rendu nous avons choisi l'autre commande *pdf2txt* mais nous nous sommes finalement retrancher après mures réflexion, la vitesse d'exécution étant plus rapide) . A la suite de ça, nous devons développer l'application permettant d'identifier Le nom du fichier d'origine (dans une ligne), Le titre du papier (dans une ligne) et Le résumé ou abstract de l'auteur (dans une ligne).

2 Choix du langage

Pour commencer nous avons fait des tests de vitesse d'exécution de la commande unix *pdftotext* en *Python* et *C++* :

Python :

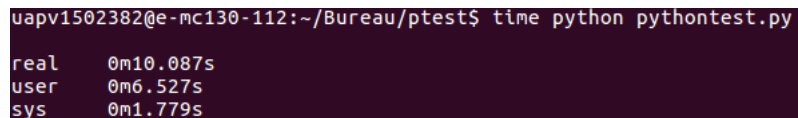
```
#!/usr/bin/env python

import subprocess
import os

n=1000;

cmd="pdftotext ../Papers/Lin_2004_Rouge.pdf"
for i in range(n) : os.system(cmd)
```

Affichage console et test a l'aide de *time* :



```
uapv1502382@e-mc130-112:~/Bureau/ptest$ time python pythontest.py
real    0m10.087s
user    0m6.527s
sys     0m1.779s
```

Figure 1. Exécution time

Ici nous avons donc une vitesse d'exécution en temps réel qui est de 10 secondes pour 1000 commande *pdftotext*. le système quand a lui a pu l'exécuter en 1 seconde.

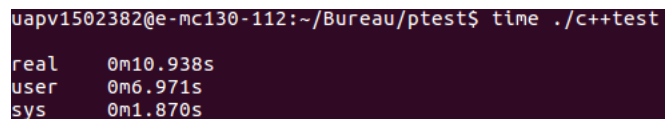
C++ :

```
#include <cstdlib>
#include <iostream>

int main(int argc, char** argv) {
    int n=1000;

    for(int i=0; i <=n ; i++){
        std::system("pdftotext ../Papers/Lin_2004_Rouge.pdf");
    }
}
```

Affichage console et test a l'aide de *time* :



```
uapv1502382@e-mc130-112:~/Bureau/ptest$ time ./c++test

real    0m10.938s
user    0m6.971s
sys     0m1.870s
```

Figure 2. Exécution time

Ici nous avons donc une vitesse d'exécution en temps réel qui est de 10 secondes pour 1000 commande *pdftotext*. le système quand a lui a pu l'exécuter en 1 seconde.

Conclusion On peut donc voir que l'exécution pour la même commande 1000 fois est la même pour les deux langage. Nous avons donc décider de choisir Python du fait que nous avons plus d'affinité avec ce langage.