

### Activité 6 : Mise en valeur du nombre de pages analysées

- Remplacez de nouveau votre fichier magazine par un extrait pour aller plus vite.

Nous allons reprendre le début du programme qui parcourt toutes les pages et initialiser une liste `liste_toutes_les_pages = []` que nous allons remplir avec toutes les url de pages rencontrées sans distinction. Ajouter les lignes en gras à votre programme.

```
fichier_magazine = open('G:\\Mes Cours\\NSI\\1ere\\Mes
cours\\projetRetroMagazineSearch\\Joystick.txt', 'r', encoding="utf-8")
lignes_magazine = fichier_magazine.readlines()

Drapeau = 0
liste_pages_resultats = []
liste_toutes_les_pages = []
for i in range(0, len(lignes_magazine)):

    if "dessinanim" in lignes_magazine[i]:
        Drapeau = 1

    if Drapeau == 1:
        if '----NOUVELLEPAGE----' in lignes_magazine[i]:
            liste_pages_resultats.append(lignes_magazine[i])
            Drapeau = 0

print(liste_pages_resultats)
print(len(liste_pages_resultats))

print((liste_toutes_les_pages))
print(len(liste_toutes_les_pages))
```

Résultat attendu : `len(liste_toutes_les_pages) = 1929`

En fin de programme, appelez votre fonction `decomptage_liens_par_dates` avec `liste_toutes_les_pages` comme argument.

On obtient :

```
1990 47 1368
1991 11 561
```

Ajouter ensuite la fréquence d'apparition du terme, c'est-à-dire le rapport (en %) entre le nombre d'occurrences du terme et le nombre de pages analysées.

```
1990 47 1368 3.435672514619883
1991 11 561 1.9607843137254901
```

*Astuce : l'affichage de texte à l'écran consomme du temps processeur, désactiver vos `print` dans la fonction `decomptage_liens_par_dates` peut démultiplier la vitesse d'exécution.*

Au final, vous éditez un fichier texte (avec le bloc note de windows ou notepad++) dans lequel on pourra trouver la liste des liens contenant le terme recherché, suivi des statistiques., en copiant les consoles de Pyzo.

```
1990 joystick_numero001
```

```
----NOUVELLEPAGE----
```

```
https://download.abandonware.org/magazines/Joystick/joystick_numero001/joystick - N°1 - janvier 1990 - page046 et page047.jpg
```

```
1990 joystick_numero002
```

```
----NOUVELLEPAGE----
```

```
https://download.abandonware.org/magazines/Joystick/joystick_numero002/joystick - N002 - fevrier 1990 - page096 et page097.jpg
```

```
1990 joystick_numero002
```

```
----NOUVELLEPAGE----
```

```
https://download.abandonware.org/magazines/Joystick/joystick_numero002/joystick - N002 - fevrier 1990 - page102 et page103.jpg
```

```
année - occurrences trouvées - total de pages analysées -  
fréquence
```

```
1990 47 1368 3.435672514619883
```

```
1991 11 561 1.9607843137254901
```

### Activité 7 :

Choisir un axe de recherche

Demande de Jean Baptiste Massuet (université Rennes 2)

*« Je pensais à une autre recherche possible, qui serait vraiment intéressante pour moi, autour des termes "digitalisation", "images digitalisées", "digitaliser"... C'était l'un des enjeux techniques à la base du photoréalisme, et à une époque, c'étaient les termes que l'on employait pour caractériser les images à tendance photographique. »*

Travail à réaliser : nous allons tout simplement chercher les occurrences de la racine « **digitalis** » dans notre script pour capter tous les termes demandés.

Magazines cibles : Joystick, Joypad, Tilt, Generation 4, Consoles +, Player One.

Remplacez maintenant votre fichier magazine (qui n'était qu'un extrait) par un fichier magazine complet et lancez l'analyse ! ATTENTION UNE OU PLUSIEURS HEURES DE CALCULS SONT NECESSAIRES !

Sauvegarder dans un fichier texte les résultats de la fenêtre de Pyzo..