

Fichiers Textes - Demo

November 26, 2019

```
[24]: f = open("demo-file-latin.txt", "r")
      #print(f.read())      # Affiche le texte en entier
      print(f.read()[:200])# Affiche les 200 premiers symboles
      f.close()
```

JULES VERNE

CINQ SEMAINES EN BALLON

VOYAGE DE DÉCOUVERTES EN AFRIQUE PAR 3 ANGLAIS

CHAPITRE PREMIER

La fin d'un discours très applaudi.--Présentation du docteur Samuel Fergusson--«
Excelsior. »

```
[29]: f = open("demo-file-latin.txt", "r")
      # affichage des longueurs de toutes les lignes
      #for line in f:
      #    print(len(line))

      # affichage des longueurs des 15 premières lignes
      c = 0
      for line in f:
          print(len(line))
          c += 1
          if c > 15:
              break
      f.close()
```

13

1

24

1

47

1

1
17
1
1
210
1
1
1
1
1
308

```
[40]: # Affichage des (début des) lignes de longueur > 800
f = open("demo-file-latin.txt", "r")
for line in f:
    if len(line) > 800:
        print(line[:75], len(line))
f.close()
```

Ils étaient là pourtant, nombreux, vieillis, fatigués, ces intrépides vo 1017
Des toasts nombreux furent portés avec les vins de France aux célèbres v 1601
Aussitôt mille encouragements se firent jour, mille félicitations éclatè 939
--Pourquoi ? répondit le docteur en s'animant; parce que jusqu'ici toutes l 802
--Non pas, mon cher Dick. J'entends bien ne pas m'en séparer avant mon arri 877
Un jour, après avoir reconnu qu'avec un bonheur insolent, on pouvait avo 890
--Eh bien ! le voici. La Société de Géographie a regardé comme très importa 893
--Enfin, reprit le docteur, M. de Heuglin, vice-consul d'Autriche à Karthou 1509
Pendant les longues heures inoccupées du voyage docteur faisait un véri 941
Il se trouvait alors à une centaine de milles de Kazeh, important établi 825
--Sans doute, mon cher Dick. Vois la marche des événements; considère les m 1766
Le missionnaire était un pauvre jeune du village d'Aradon, en Bretagne, 1834
Le docteur ne put rien obtenir d'eux, et il partit seul au milieu de la 977
--Volontiers. James Bruce était un Écossais du comté de Stirling, qui, de 1 1429
--Oui, cette contrée est fatale ! Nous marchons directement vers le royaume 1533
Que de fois ses regards se portèrent en l'air ! Il espérait apercevoir l 887
« Je me croyais perdu, mon maître, dit-il, et mes pensées s'adressaient 2649
--Il y a longtemps, répondit le docteur. La reconnaissance du Niger et de s 1414
--C'est alors qu'un courageux jeune homme entreprit avec ses faibles ressou 1728
--D'ailleurs, si je ne me trompe, mes amis, le vent a une légère tendance à 808
--Le voici. En 1854, un marabout du Fouta sénégalais, Al-Hadji, se disant i 1099

```
[32]: # Idiom python standard pour la manipulation des fichiers
liste_lignes = []
with open("demo-file-utf8.txt", "r", encoding="utf8") as f:
    for ligne in f:
        liste_lignes.append(ligne.strip())
# Maintenant la liste "liste_lignes" contient la totalité des lignes du fichier
↳ (sauf retours-chariot)
```

```
[33]: # Nombre de lignes du fichier
# On suppose le fichier déjà lu et copié dans liste_lignes
print("Le fichier comprend ",len(liste_lignes), "lignes.")
# Nombre de caractères du fichier
nb_car = 0
for l in liste_lignes:
    nb_car = nb_car + len(l)
print("Le fichier comprend %d caractères." % nb_car)
```

Le fichier comprend 5823 lignes.
Le fichier comprend 486359 caractères.

```
[34]: # Mêmes opérations sans passer par une liste en mémoire
nb_lgn = 0
nb_car = 0
with open("demo-file-utf8.txt", "r", encoding="utf8") as f:
    for l in f:
        nb_lgn += 1
        nb_car += len(l.strip())
print("Le fichier comprend %d lignes et %d caractères." % (nb_lgn, nb_car))
```

Le fichier comprend 5823 lignes et 486359 caractères.

```
[35]: # Affichage de la plus longue ligne de la liste
# On suppose le fichier déjà lu et copié dans liste_lignes
l_max = 0
ligne_l_max = ""
for l in liste_lignes:
    if len(l) > l_max:
        ligne_l_max = l
        l_max = len(l)
print("La plus longue ligne comprend %d caractères et commence par //%s//" % (
    ↪(l_max, ligne_l_max[:15]))
```

La plus longue ligne comprend 2645 caractères et commence par //« Je me croyais//

```
[39]: # Affichage (du début) des lignes qui commencent par "Le"
with open("demo-file-utf8.txt", "r", encoding="utf8") as f:
    for l in f:
        if l.startswith("Le"):
            print(l[:75])
```

Le docteur Fergusson avait un ami. Non pas un autre lui-même, un alter ego;
Le 16 février, le Resolute vint jeter l'ancre devant Greenwich. C'était un
Le docteur Fergusson termina ainsi son discours, et fut applaudi de bon c u
Le docteur fut alors obligé de soutenir son aérostat par une dilatation de
Les montagnes de la Lune.--Un océan de verdure.--On jette l'ancre.--L'éléph

Le Karagwah.--Le lac Ukéréoué.--Une nuit dans une île.--L'Équateur.--Traver
Le docteur regarda une dernière fois ces précieuses initiales, dont il prit
Le Nil.--La Montagne tremblante.--Souvenir du pays.--Les récits des Arahes.
Les sacs étaient disposés, les armes étaient en état.

Le vent tombe.--Les approches du Désert.--Le décompte de la provision d'eau
Le pays dans le coude du Niger.--Vue fantastique des monts Hombori.--Kabra.
Les approches du Sénégal.--Le Victoria baisse de plus en plus.--On jette, o
Les Talibas.--La poursuite.--Un pays dévasté.--Vent modéré.--Le Victoria ba

```
[44]: # Affichage des lignes qui contiennent le motif "ménage"
with open("demo-file-utf8.txt", "r", encoding="utf8") as f:
    for l in f:
        if l.find("ménage") != -1:
            print(l[:75])
```

Importance de Joe. --Le commandant de la Resolute.--L'arsenal de Kennedy.--
--Si vraiment ; il faudra d'ailleurs ménager nos vivres, et, chemin faisant
Le Victoria faisait peu de chemin, et le vent semblait vouloir ménager s
C'était toute une ménagerie rare dans une serre merveilleuse, où des ois
--Cet animal n'est vulnérable qu'au ventre et entre les cuisses; la balle d
--Du gaz, mon garçon ; la force ascensionnelle du ballon diminue sensiblement

```
[67]: # Affichage des lignes qui contiennent "esca" avec mise en relief
with open("demo-file-utf8.txt", "r", encoding="utf8") as f:
    for l in f:
        if l.find("esca") != -1:
            decoup = l.partition("esca")
            print ("%20s %s %s" % (decoup[0][-20:-1], decoup[1], decoup[2][:
↳20]))
```

Il esca lada la nacelle et d
nc ? demanda Joe en esca ladant la nacelle.

es avaient en effet esca ladé le baobab; ils
l'intérieur par un esca lier qui n'était plu
issante roula sur l esca lier, renversant Joe
his les halliers, j esca lade les palissades,

```
[38]: # Affichage de tous les mots qui commencent par une majuscule
# Il faut "découper" les lignes avec split()
# pour afficher les 20 premiers on utilise un compteur et l'instruction _break_
with open("demo-file-utf8.txt", "r", encoding="utf8") as f:
    c = 0
    for l in f:
        mots = l.split()
        for m in mots:
            if m[0].isupper():
```

```

        print(m)
        c += 1
    if c > 20:
        break

```

VERNE
 CINQ
 SEMAINES
 EN
 BALLON
 VOYAGE
 DE
 DÉCOUVERTES
 EN
 AFRIQUE
 PAR
 ANGLAIS
 CHAPITRE
 PREMIER
 La
 Samuel
 Fergusson--«
 Excelsior.
 Traveller's
 Il
 Société
 Londres,
 Waterloo
 Le
 Francis
 M...

```

[22]: # Fréquence de tous les mots qui commencent par une majuscule
      dico = {}
      with open("demo-file-utf8.txt", "r", encoding="utf8") as f:
          for l in f:
              mots = l.split()
              for m in mots:
                  if m[0].isupper():
                      dico[m] = dico.get(m,0) + 1
      # print(dico)
      for key in sorted(dico)[:10]:
          print("le mot %s apparaît %d fois." % (key,dico[key]))

```

le mot A apparaît 85 fois.
 le mot A. apparaît 2 fois.
 le mot AFRIQUE apparaît 1 fois.
 le mot ANGLAIS apparaît 1 fois.

le mot Abbadie, apparaît 1 fois.
le mot Aborder apparaît 1 fois.
le mot Abyssiniens apparaît 1 fois.
le mot Accrochons-nous apparaît 1 fois.
le mot Achille apparaît 1 fois.
le mot Adam, apparaît 1 fois.

[]: #