

```
In [1]: import json
import time
from bs4 import BeautifulSoup
import os
from tqdm import tqdm

In [2]: data_dir = os.path.join("../blogger_blogs")
blog_dirs = os.listdir(data_dir)

In [3]: liste_posts = []
data_M = []

for i in range(len(blog_dirs)) :
    blog_id = blog_dirs[i]
    with open(os.path.join( data_dir, blog_id, "blog_posts%s.json" % blog_id), 'r') as f :
        data = json.load(f)
        data_M = data_M + data
        liste_posts.append(data)

In [4]: len(liste_posts)

Out[4]: 187

In [5]: print(type(data))
print(len(data))
#pourquoi ??

<class 'list'>
63

In [6]: print(type(data_M))
print(len(data_M))

<class 'list'>
110895

In [7]: type(liste_posts)

Out[7]: list

In [8]: with open(os.path.join('posts.json'), 'w', encoding = 'UTF-8') as fin :
    json.dump( liste_posts, fin)

In [9]: with open(os.path.join('posts_maior.json'), 'w', encoding = 'UTF-8') as fin_M :
    json.dump( data_M, fin_M)

In [10]: print("Auteur : ", data_M[0]['author']['displayName'], "\n", data_M[0]['content'])

Auteur : Heptanes Fraxion
<p><span style="font-family: courier; font-size: large;"><br /></span></p><p><span style="font-family: courie
r; font-size: large;"><br /></span></p><p><span style="font-family: courier; font-size: large;">le premier jour du
premier mois de l'année ressemble étrangement au dernier jour du dernier mois de l'année&nbsp;</span></p><p>
<span style="font-family: courier; font-size: large;">plus rien ne distingue l'aube du crépuscule&nbsp;</span>
</p><p><span style="font-family: courier; font-size: large;">et toi qui flottes à mes côtés lorsque que tu me r
egardes ainsi j'ai l'impression d'être sauvé&nbsp;</span></p><p><span style="font-family: courier; font-size: l
arge;">il neige des noms et jamais je n'en retiens un seul&nbsp;</span></p><p><span style="font-family: courier; font-
size: large;"><br /></span></p><p><span style="font-family: courier; font-size: large;">sous ce ciel là&nbsp;<
/ps><p><span style="font-family: courier; font-size: large;">la terre qui rêve et qui tourne semble soud
ain souteraine&nbsp;</span></p><p><span style="font-family: courier; font-size: large;">bournes de bitume dan
s les odeurs saines de l'hiver&nbsp;</span></p><p><span style="font-family: courier; font-size: large;">j'enchain
e des dizaines de patelins&nbsp;</span></p><p><span style="font-family: courier; font-size: large;">)<état a u
employeur qu'in&nbsp;</span></p><p><span style="font-family: courier; font-size: large;">créé des poisons avant que de vendre ses antidotes et toi te l'ai-j
e déjà dit à peine me touches-tu que mes cicatrices disparaissent&nbsp;</span></p><p><span style="font-family: courie
r; font-size: large;"><br /></span></p><p><span style="font-family: courier; font-size: large;">splendeur sérénité
sandwich thon crudités sainte trinité qui me rend immortel sur l'aire de repos&nbsp;</span></p><p><span style=
"font-family: courier; font-size: large;">tel le poète maudit de base</span></p><p><span style="font-family: c
ourier; font-size: large;">je porte un bonnet post apo et comme n'importe quel ado déséquilibré je me suis offe
rt un couteau de poche</span></p><p><span style="font-family: courier; font-size: large;">et toi que je cherche
depuis tant de temps je m'aperçois qu'en fait tu as toujours été là&nbsp;</span></p><p><span style="font-famil
y: courier; font-size: large;">il neige des noms et jamais je n'en retiens un seul&nbsp;</span></p></p>

In [11]: blog_dirs_minor = blog_dirs[0:5]
len(blog_dirs_minor)

liste_posts_minor = []
data_m = []

for i in range(len(blog_dirs_minor)) :
    blog_id = blog_dirs_minor[i]
    with open(os.path.join( data_dir, blog_id, "blog_posts%s.json" % blog_id), 'r') as f2 :
        blog_posts_minor = json.load(f2)
        data_m = data_m + blog_posts_minor
        liste_posts_minor.append(blog_posts_minor)

with open(os.path.join('posts_minor.json'), 'w', encoding = 'UTF-8') as fin2 :
    json.dump( liste_posts_minor, fin2)

In [12]: data_m[0]['content']

Out[12]: '<p><span style="font-family: courier; font-size: large;"><br /></span></p><p><span style="font-family: courie
r; font-size: large;"><br /></span></p><p><span style="font-family: courier; font-size: large;">le premier jour du
premier mois de l'année ressemble étrangement au dernier jour du dernier mois de l'année&nbsp;</span></p><p>
<span style="font-family: courier; font-size: large;">plus rien ne distingue l'aube du crépuscule&nbsp;</spa
n></p><p><span style="font-family: courier; font-size: large;">et toi qui flottes à mes côtés lorsque que tu me r
egardes ainsi j'ai l'impression d'être sauvé&nbsp;</span></p><p><span style="font-family: courier; font-siz
e: large;">il neige des noms et jamais je n'en retiens un seul&nbsp;</span></p><p><span style="font-family: courie
r; font-size: large;"><br /></span></p><p><span style="font-family: courier; font-size: large;">sous ce ciel là&
nbsp;</span></p><p><span style="font-family: courier; font-size: large;">la terre qui rêve et qui tourne semble s
oudain souteraine&nbsp;</span></p><p><span style="font-family: courier; font-size: large;">bournes de bitume da
ns les odeurs saines de l'hiver&nbsp;</span></p><p><span style="font-family: courier; font-size: large;">)<état a
un employeur qu'in&nbsp;</span></p><p><span style="font-family: courier; font-size: large;">créé des poisons avant que de vendre ses antidotes et toi te l'ai-j
e déjà dit à pei
ne me touches-tu que mes cicatrices disparaissent&nbsp;</span></p><p><span style="font-family: courier; font-siz
e: large;"><br /></span></p><p><span style="font-family: courier; font-size: large;">splendeur sérénité
sandwich thon crudités sainte trinité qui me rend immortel sur l'aire de repos&nbsp;</span></p><p><span style="fo
nt-family: courier; font-size: large;">tel le poète maudit de base</span></p><p><span style="font-famil
y: courier; font-size: large;">je porte un bonnet post apo et comme n'importe quel ado déséquilibré je me
suis offert un couteau de poche</span></p><p><span style="font-family: courier; font-size: large;">et toi que j
e cherche depuis tant de temps je m'aperçois qu'en fait tu as toujours été là&nbsp;</span></p><p><span style=
"font-family: courier; font-size: large;">il neige des noms et jamais je n'en retiens un seul&nbsp;</span></p>
>'

In [13]: len(data_m)

Out[13]: 1378

In [14]: len(data_M)

Out[14]: 110895

In [ ]:

In [20]: data_M[95873]

Out[20]: {'kind': 'blogger#post',
'id': '3665579785497953424',
'blog': {'id': '7310156134283777261'},
'published': '2015-08-03T00:15:00+02:00',
'updated': '2015-08-03T00:19:21+02:00',
'url': 'http://feusurlequartiergeneral.blogspot.com/2015/08/ramuz-ecrivain-panique.html',
'selfLink': 'https://www.googleapis.com/blogger/v3/blogs/7310156134283777261/posts/3665579785497953424',
'title': 'Ramuz, écrivain panique.',
'content': '<div dir="ltr" style="text-align: left;" trbidi="on">\n<div class="separator" style="clear: both;
text-align: center;">\n<a href="http://3.bp.blogspot.com/-3fPEefVhRlI/Vb6WSAK_9pI/AAAAAAAAUcM/KIuQtW9sa28/s160
0/R150127827.jpg" imageanchor="1" style="margin-left: 1em; margin-right: 1em;"><img border="0" height="640" src=
"http://3.bp.blogspot.com/-3fPEefVhRlI/Vb6WSAK_9pI/AAAAAAAAUcM/KIuQtW9sa28/s640/R150127827.jpg" width="403" />
</a></div>\n<div style="text-align: justify;">\n<br />\n<b>&nbsp;</b><i>paru sur Causeur.fr</i><br />\n<br />\n
En France, on ne lit pas Ramuz parce qu'on a Jean Giono. C'est \ninjuste mais c'est comme ça: Giono est fran
çais, Ramuz est suisse. On a \ntoujours tort, malgré tout, de se priver d'un grand écrivain. On \nobjectera que
Ramuz est depuis 2005 édité en Pléiade en deux volumes. \nC'est parce que les autorités helvètes ont mis la mai
n à la poche et, de\n toute manière, la Pléiade n'est pas forcément la garantie de lecteurs \nnouveaux: on ne
découvre pas un écrivain en Pléiade, on le conserve. On \nle conserve parce qu'on a aimé ses livres sous une aut
re forme avant et \nque le papier bible nous fait gagner de la place et s'abime moins vite. \nMais allez trouve
r les romans de Ramuz en poche... C'est pratiquement \nmission impossible y compris pour son titre le plus connu:
<i>La grande peur dans la montagne</i>. </b></div>\n<b>\n</b></div style="text-align: justify;">\n<b>Ramuz, Charles Ferdinand de son prénom, est né en 1878 à Lausanne et \nmort à Pull y en 1947. On voit qu'il ne fit que lue
ques kilomètres dans \nsa vie. La Suisse lui a suffi, et dans la Suisse, le canton de Vaud. Les\ngands écrivai
ns n'ont pas forcément besoin de voyager. L'infini est \nsur le pas de leur porte, il suffit de trouver l'\angl
e et la lumière \npour en donner toute la mesure. Ramuz, c'est l'anti-Cendrars, \nl'anti-Morand, l'anti-Larbaud
qui à la même époque voyageait beaucoup \ndans les cargos, les steamers, les avions, les trains express au
x \n&nbsp;<i>bruits miraculeux&nbsp;</i>. </b></div>\n<b>\n</b></div style="text-align: justify;">\n<b>À part u
ne période parisienne qu'il interromp en 1914 parce que ce \nqu'il sent venir l'horrible, Ramuz n'aura pratiq
ument pas bougé du \ncanton de Vaud. Il a eu raison. Rien n'est plus exotique que le canton \nde Vaud pour peu
que vous ayez un style pour le dire&nbsp;<!-- ce qui compte, \nc'est le prisme, pas le décor. Et le style de Ramuz
rend étrange, épique\n ou tragique ce qu'il y a de plus banal. Nous ne disons pas que le \ncanton de Vaud est b
anal, nous ne voudrions pas nous attirer les foudres\nde Roland Jaccard mais tout de même, le lecteur est touj
ours un \npeu surpris en lisant <i>La grande peur dans la montagne</i> d'avoir \nl'impression d'être dans un mon
de aussi brutal, primitif, plein de \nsortilèges et de violences que chez Faulkner ou Giono. Mais on sait \nausi
si, justement, que le Sud de Faulkner ou la Provence de Giono sont \nlargement oniriques, fantasmatiques, les \nco
uleurs inédites, la description d'une nature panique, c'est cette \nfausse précision des rêves et des hallucinations psy
chotiques.</b></div>\n<b>\n</b></div style="text-align: justify;">\n<b><i>La grande peur dans la montagne</i> p
eut d'ailleurs être lu \ncomme le roman d'une psychose si l'on veut rester cartésien ou comme un \ntexte fanta
stique si on garde une âme d'enfant. Depuis \nTodorov, il est clair que le fantastique, c'est précisément l'hési
tation, \nl'incertitude et que c'est de cette hésitation, de cette incertitude que \nnaît notre malaise, voire
notre angoisse. On a beaucoup de mal à \ncomprendre ça, dans le
```