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]: In [5]: print(type(data)) print(len(data)) #pourquoi ?? <class 'list'> In [6]: print(type(data M)) print(len(data M)) <class 'list'> 110895 In [7]: type(liste posts) Out[7]: with open(os.path.join('posts.json'), 'w', encoding = 'UTF-8') as fin : In [8]: 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 <span style="font-family: courier; font-size: large;"><br /></span><span style="font-family: courie r; font-size: large;"><br /></span><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 </span> <span style="font-family: courier; font-size: large;">plus rien ne distingue l'aube du crépuscule&nbsp;</span> <span style="font-family: courier; font-size: large;">et toi qui flottes à mes côtés lorsque que tu me r  $egardes \ ainsi \ j'ai \ l'impression \ d'être \ sauv\'e\ </span><span \ style="font-family: courier; font-size: left to be a size of the left to be a size of th$ arge;">il neige des noms et jamais je n'en retiens un seul</span><span style="font-family: courier; font -size: large;"><br /></span><span style="font-family: courier; font-size: large;">sous ce ciel là&nbsp; </span><span style="font-family: courier; font-size: large;">la terre qui rêve et qui tourne semble soud ain souterraine </span><span style="font-family: courier; font-size: large;">bornes de bitume dans les odeurs saines de l'hiver </span><span style="font-family: courier; font-size: large;">j'enchaîn e des dizaines de patelins </span><span style="font-family: courier; font-size: large;">l'état a un employeur qui créé des poisons avant que de vendre ses antidotes et toi te l'ai-je déjà dit à peine me to uches-tu que mes cicatrices disparaissent </span><span style="font-family: courier; font-size: larg e;">il neige des noms et jamais je n'en retiens un seul </span><span style="font-family: courier; f ont-size: large;"><br /></span><span style="font-family: courier; font-size: large;">splendeur sérénité sandwich thon crudités sainte trinité qui me rend immortel sur l'aire de repos </span><span style ="font-family: courier; font-size: large;">tel le poète maudit de base</span><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><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 </span><span style="font-famil y: courier; font-size: large;">il neige des noms et jamais je n'en retiens un seul </span> 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'] '<span style="font-family: courier; font-size: large;"><br /></span><span style="font-family: courie Out[12]: r; font-size: large;"><br /></span><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 </span> p><span style="font-family: courier; font-size: large;">plus rien ne distingue 1\'aube du crépuscule&nbsp;</spa n><span style="font-family: courier; font-size: large;">et toi qui flottes à mes côtés lorsque que tu me regardes ainsi j\'ai l\'impression d\'être sauvé </span><span style="font-family: courier; font-siz e: large;">il neige des noms et jamais je n\'en retiens un seul</span><span style="font-family: courier; font-size: large; "><br /></span><span style="font-family: courier; font-size: large;">sous ce ciel là&nb sp;</span><span style="font-family: courier; font-size: large;">la terre qui rêve et qui tourne semble s oudain souterraine </span><span style="font-family: courier; font-size: large;">bornes de bitume da ns les odeurs saines de l\'hiver </span><span style="font-family: courier; font-size: large;">j\'en chaîne des dizaines de patelins </span><span style="font-family: courier; font-size: large;">1\'éta t a un employeur qui créé des poisons avant que de vendre ses antidotes et toi te l\'ai-je déjà dit à pei ne me touches-tu que mes cicatrices disparaissent </span><span style="font-family: courier; font-si ze: large;">il neige des noms et jamais je n\'en retiens un seul </span><span style="font-family: c ourier; font-size: large;"><br /></span><span style="font-family: courier; font-size: large;">splendeur sérénité sandwich thon crudités sainte trinité qui me rend immortel sur l\'aire de repos </span><sp an style="font-family: courier; font-size: large;">tel le poète maudit de base</span><span style="font-f amily: 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><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 </span><span style ="font-family: courier; font-size: large;">il neige des noms et jamais je n\'en retiens un seul </span></p len(data m) In [13]: Out[13]: In [14]: len(data M) 110895 Out[14]: In [ ]: In [20]: data\_M[95873] {'kind': 'blogger#post', Out[20]: '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/-3fPEefVhR1I/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/-3fPEefVhR1I/Vb6WSAk\_9pI/AAAAAAAAUCM/KIuQtW9sa28/s640/R150127827.jpg" width="403" />  $</a></div>\n<div style="text-align: justify;">\n<br/>>\n<br/>>\n<br/>>bsp;</b><i>paru sur Causeur.fr</i><br/>>\n<br/>br />\n<br/>$ <b>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 d écouvre 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>\n</b>\n</b>\n<div style="text-align: justify;">\n<b>Ramuz, Cha rles Ferdinand de son prénom, est né en 1878 à Lausanne et \nmort à Pully en 1947. On voit qu'il ne fit que que lques kilomètres dans \nsa vie. La Suisse lui a suffi, et dans la Suisse, le canton de Vaud. Les\n grands écriv ains n'ont pas forcément besoin de voyager. L'infini est \nsur le pas de leur porte, il suffit de trouver l'ang le et la lumière \npour en donner toute la mesure. Ramuz, c'est l'anti-Cendrars, \nl'anti-Morand, l'anti-Larbau d qui à la même époque voyagent beaucoup \ndans les cargos, les steamers, les aéroplanes, les trains express au x « <i>bruits miraculeux&nbsp;</i>».</b>\n<b>\n<div style="text-align: justify;">\n<b>À part u ne période parisienne qu'il interrompt en 1914 parce que ce \nqu'il sent venir l'horrifie, Ramuz n'aura pratiqu ement 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 : 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\n de Roland Jaccard mais tout de même, le lecteur est touj ours un \npeu surpris en lisant <i>La grand 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 \naus si, justement, que le Sud de Faulkner ou la Provence de Giono sont \nlargement oniriques, fantasmatiques et que la précision avec laquelle \nils nous sont rendus au travers de l'épaisseur des personnages, les \ncouleurs iné dites, la description d'une nature panique, c'est cette \nfausse précision des rêves et des hallucinations psyc hotiques.</b></div>\n<b>\n<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 fantas tique 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 ca, dans le fantastique contemporain où l'existence des \nf antômes nous est donnée comme allant de soi dans les pénibles romans de\n Marc Levy, par exemple.</b></div>\n<b >\n</b>\n<div style="text-align: justify;">\n<b><i>La grande peur dans la montage</i>, qui refuse &nbsp;le temp s et \nl'espace, peut néanmoins par déduction être situé dans un village de \nmontagne reculé dans les années v ingt. C'est sûrement un village suisse \ncar les personnages disent « septante » et qu'il n'y a pas d e montagnes \nen Belgique.</b></div>\n<b>\n</div style="text-align: justify;">\n<b>Là comme partout et de tout temps, les vieux s'opposent aux jeunes. \nEn l'occurrence, à propos d'un pâturage. Ce pâturage, à 2300 mèt res d'altitude, \nreste inutilisé depuis vingt ans. Et ça ne plait pas au nouveau maire \nqui est du parti des jeunes. Les vieux, eux, préfèrent laisser ce \ncoin-là tranquille et tant pis pour les pertes engendrées pour l a \ncommune: il s'est passé là haut une tragédie. Une force mystérieuse a \ndécimé les hommes qui y passaient l es deux mois d'été à faire paître les\n vaches et à fabriquer du lait. Par une belle majorité au conseil \ngéné ral qui est, si on a bien compris, la réunion de tous les habitants \net pas seulement du conseil municipal, ce qui tendrait à prouver qu'un \nvillage suisse des années vingt a un fonctionnement qui le rapproche \ndavantage d'Athènes au Vème siècle que de Bruxelles au XXIème, il est \ndonc décidé d'en finir avec les superstitions anc iennes. </b></div>\n<div style="text-align: justify;">\n<b>On monte une \nexpédition pour réparer le chale t où sera fait le fromage et puis sept \nvolontaires se rendent là haut. Ils ont des intentions diverses. Il y a un \namoureux qui veut avoir de quoi épouser sa belle, un garçon très laid \nqui veut trouver de l'or (chez R amuz comme dans l'Antiquité, l'habit \nfait le moine), un vieux qui était déjà là vingt ans plus tôt mais qui \nse croit protégé par un verset biblique trempé trois fois dans l'eau \nd'un lac.</b>\n/b>\n/b>\n/div s tyle="text-align: justify;">\n<b>À partir de ce moment, <i>La grande peur dans la montagne</i> prend l'allure d e <i>Dix petits nègres</i>\n dans les alpages. Les morts suspectes, les coups du sort se succèdent. A\n chaque fois, le rationaliste veut trouver une explication rationnelle. \nIl y arrive de moins en moins. Ramuz a remarq uablement distillé la peur, \n notamment dans la spatialisation du roman: en bas le village, en haut \nl'alpage, au milieu un chemin compliqué entre forêts, torrents, \nglaciers. Et toute la peur vient aussi de ce qu'il ne f aut absolument pas \nque le haut contamine le bas, et on parle ici de contamination au\n sens propre puisque le bétail commence à crever.</b></div>\n<b>\n<div style="text-align: justify;">\n<b>On est étonné, à la fin, par la force tellurique de ce \nbref roman. Même l'écriture de Ramuz, si critiquée en son temps pour son \n mélange de trivialité réaliste, de répétitions bibliques et de syntaxe \nhomérique, finit par séduire ou au moins par créer chez le lecteur qui \nveut bien se laisser faire, cet envoûtement hypnotique qui est ce qu'on \ndemande d'abord à la littérature.</b>\n<b>\n</div>\n<div style="text-align: justify;">\n<br/>br /></div>\n<di v style="text-align: justify;">\n<br /></div>\n<div style="text-align: justify;">\n<br />></div>\n<div style="text-align: justify;">\n<div style="tex xt-align: justify;">\n<b><i>La Grande Peur dans la Montagne</i> de Charles Ferdinand Ramuz (<i>Livre de poche, édition 1972, 30 centimes, vide-grenier du quartier Tujac à Brive</i>)</b></div>\n<br/>br /></div>\n', 'author': {'id': '12658941614607286284', 'displayName': 'Jérôme Leroy', 'url': 'https://www.blogger.com/profile/12658941614607286284', 'image': {'url': '//4.bp.blogspot.com/-tf7LQ9wnSZs/YD57ySG-O2I/AAAAAAAbbk/1-DrKomoLoI6kkZZck LtBYllwQJ4C7CgC K4BGAYYCw/s35/Je%252525CC%25252581ro%252525CC%25252582me Leroy %252525282018%25252529.jpg'}}, 'replies': {'totalItems': '1', 'selfLink': 'https://www.googleapis.com/blogger/v3/blogs/7310156134283777261/posts/3665579785497953424/commen 'labels': ['mes vacances chez les bouquinistes', 'ramuz', 'suisse'], 'etag': '"dGltZXN0YW1w0iAxNDM4NTUzOTYxMzc0Cm9mZnNldDogNzIwMDAwMAo"', 'ext': ['\n', '\n', '\n', '\n', '\n', '\xa0', 'paru sur Causeur.fr', '\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 d écouvre 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: 'La grande peur dans la montagne', '\n', '\n', '\n', '\n', 'Ramuz, Charles Ferdinand de son prénom, est né en 1878 à Lausanne et \nmort à Pully en 1947. On voit qu'il n e fit que quelques kilomètres dans \nsa vie. La Suisse lui a suffi, et dans la Suisse, le canton de Vaud. Les\n grands écrivains n'ont pas forcément besoin de voyager. L'infini est \nsur le pas de leur porte, il suffit de t rouver l'angle 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 voyagent beaucoup \ndans les cargos, les steamers, les aéroplanes, les trai ns express aux «\xa0', 'bruits miraculeux\xa0', '».', '\n', '\n', '\n', '\n', 'À part une période parisienne qu'il interrompt en 1914 parce que ce \nqu'il sent venir l'horrifie, Ramuz n'a ura pratiquement pas bougé du \ncanton de Vaud. Il a eu raison. Rien n'est plus exotique que le canton \nde Vau d pour peu que vous ayez un style pour le dire\xa0: 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 V aud est banal, nous ne voudrions pas nous attirer les foudres\n de Roland Jaccard mais tout de même, le lecteur est toujours un \npeu surpris en lisant ', 'La grand peur dans la montagne', ' d'avoir \nl'impression d'être dans un monde aussi brutal, primitif, plein de \nsortilèges et de violences q ue chez Faulkner ou Giono. Mais on sait \naussi, justement, que le Sud de Faulkner ou la Provence de Giono sont \nlargement oniriques, fantasmatiques et que la précision avec laquelle \nils nous sont rendus au travers de l'épaisseur des personnages, les \ncouleurs inédites, la description d'une nature panique, c'est cette \nfausse précision des rêves et des hallucinations psychotiques.', '\n', '\n', '\n', '\n', 'La grande peur dans la montagne', ' peut d'ailleurs être lu \ncomme le roman d'une psychose si l'on veut rester cartésien ou comme un \ntexte f antastique si on garde une âme d'enfant. Depuis \nTodorov, il est clair que le fantastique, c'est précisément l'hésitation, \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 fantastique contemporain où l'existence d es \nfantômes nous est donnée comme allant de soi dans les pénibles romans de\n Marc Levy, par exemple.', '\n', '\n', '\n', '\n', 'La grande peur dans la montage', ', qui refuse \xa0le temps et \nl'espace, peut néanmoins par déduction être situé dans un village de \nmontag ne reculé dans les années vingt. C'est sûrement un village suisse \ncar les personnages disent «\xa0septante\xa 0» et qu'il n'y a pas de montagnes \nen Belgique.', '\n', '\n', '\n', '\n', "Là comme partout et de tout temps, les vieux s'opposent aux jeunes. \nEn l'occurrence, à propos d'un pâturag e. Ce pâturage, à 2300 mètres d'altitude, \nreste inutilisé depuis vingt ans. Et ça ne plait pas au nouveau mai re \nqui est du parti des jeunes. Les vieux, eux, préfèrent laisser ce \ncoin-là tranquille et tant pis pour le s pertes engendrées pour la \ncommune: il s'est passé là haut une tragédie. Une force mystérieuse a \ndécimé le s hommes qui y passaient les deux mois d'été à faire paître les\n vaches et à fabriquer du lait. Par une belle majorité au conseil \ngénéral qui est, si on a bien compris, la réunion de tous les habitants \net pas seulemen t du conseil municipal, ce qui tendrait à prouver qu'un \nvillage suisse des années vingt a un fonctionnement q ui le rapproche \ndavantage d'Athènes au Vème siècle que de Bruxelles au XXIème, il est \ndonc décidé d'en fini r avec les superstitions anciennes.\xa0", '\n', 'On monte une \nexpédition pour réparer le chalet où sera fait le fromage et puis sept \nvolontaires se rende nt là haut. Ils ont des intentions diverses. Il y a un \namoureux qui veut avoir de quoi épouser sa belle, un g arçon très laid \nqui veut trouver de l'or (chez Ramuz comme dans l'Antiquité, l'habit \nfait le moine), un vie ux qui était déjà là vingt ans plus tôt mais qui \nse croit protégé par un verset biblique trempé trois fois da ns l'eau \nd'un lac. '\n', '\n', '\n', '\n', 'À partir de ce moment, ', 'La grande peur dans la montagne', ' prend l'allure de ', 'Dix petits nègres', '\n dans les alpages. Les morts suspectes, les coups du sort se succèdent. A\n chaque fois, le rationaliste v eut trouver une explication rationnelle. \nIl y arrive de moins en moins. Ramuz a remarquablement distillé la p eur, \n notamment dans la spatialisation du roman: en bas le village, en haut \nl'alpage, au milieu un chemin co mpliqué entre forêts, torrents, \nglaciers. Et toute la peur vient aussi de ce qu'il ne faut absolument pas \nq ue le haut contamine le bas, et on parle ici de contamination au\n sens propre puisque le bétail commence à cre ver.', '\n', '\n', '\n', '\n', 'On est étonné, à la fin, par la force\xa0 tellurique de ce \nbref roman. Même l'écriture de Ramuz, si critiq uée en son temps pour son\n mélange de trivialité réaliste, de répétitions bibliques et de syntaxe \nhomérique, finit par séduire ou au moins par créer chez le lecteur qui \nveut bien se laisser faire, cet envoûtement hypno tique qui est ce qu'on \ndemande d'abord à la littérature.', '\n', 'La Grande Peur dans la Montagne', ' de Charles Ferdinand Ramuz (', 'Livre de poche, édition 1972, 30 centimes, vide-grenier du quartier Tujac à Brive', ')', '\n', '\n']} In [ ]: In [ ]: | z = '' list clean content = [] for j in range(len(data M)) : index = data M[j]['content'] #index1 = "Auteur : " , data M[0]['author']['displayName'], "\n" soup = BeautifulSoup(index.replace("<br />", "\n")) z = data M[j]['author']['displayName'], "\n" , soup.get text().replace("\xa0", " ") list clean content.append(z) In [ ]: len(list clean content) In [ ]: list clean content[0] In [ ]: list clean content[1377] In []: with open(os.path.join('posts maior propre.json'), 'w', encoding = 'UTF-8') as fin M2: json.dump( list clean content, fin M2) In [ ]: import pandas as pd df = pd.read\_json (r"posts\_maior\_propre.json") df.to\_csv (r"post\_contents.txt", index = False)