



Script Anais-Anpuh

Link: <https://github.com/ericbrasiln/Anais-Anpuh>

Projeto de script para web scraping da página de [Anais da Associação Nacional de História – Anpuh](#). Parte de projeto de História Digital desenvolvido no colegiado do curso de Licenciatura em História (CCLHM) da Unilab, campus dos Malês, sob coordenação do professor [Eric Brasil](#) em parceria com o [Laboratório de Humanidades Digitais da Ufba](#).

O Script Anais-Anpuh realiza a raspagem dos papers em pdf de todos os Simpósios Nacionais da Anpuh entre 1963 até 2017 (disponíveis atualmente na site).

A ferramenta foi desenvolvida apenas para pesquisas acadêmicas, sem fins lucrativos.

Esse script foi pensado como uma ferramenta metodológica da pesquisa em humanidades digitais. Sua criação é fruto das reflexões e experiências empíricas de historiadores e sociólogos que têm enfrentado o [desafio de fazer ciências humanas no mundo digital](#).

Defendemos a importância da apropriação, uso, desenvolvimento e aprimoramento de ferramentas digitais para as humanidades, assim como a urgência na sofisticação teórica, metodológica e epistemológica sobre as chamadas Humanidades Digitais.

É crescente o número de repositórios de fontes e dados on-line, assim como o acesso, busca, pesquisa e, muitas vezes, dependência de pesquisadores/as a eles.

Os Simpósios Nacionais da Anpuh, que acontecem bienalmente, têm reunido importantes reflexões sobre as mais variadas perspectivas historiográficas. Por conseguinte, os anais de cada evento constituem um importante repositório para pesquisas nos mais variados campos de estudo.

Índice

- [Script Anais-Anpuh](#)
- [Índice](#)
- [Instalação](#)
 - [Python](#)
 - [Bibliotecas e módulos](#)
- [Resultados](#)
- [Licença](#)

Instalação

Para executar o Script Anais-Anpuh, vc precisa acessar a pasta da ferramenta no [GitHub](#). Clone ou faça download do repositório e salve na pasta que deseja que os resultados e seus respectivos arquivos sejam armazenados. Antes de executar o script, é preciso preparar seu computador, como mostramos abaixo.

Python

A ferramenta consiste num script escrito em [Python 3.8](#). Esta é uma linguagem de programação que te permite trabalhar rapidamente e integrar diferentes sistemas com maior eficiência.

Para executar o arquivo .py é preciso instalar o Python3 em seu computador.

[Clique aqui](#) para um tutorial de instalação do Python no Windows, [clique aqui](#) para Linux e [clique aqui](#) para Mac.

Após a instalação, vc pode executar o arquivo .py direto do prompt de comando do Windows ou pelo terminal do Linux, ou utilizar as diversas [IDE](#) disponíveis.

Exemplo de como executar utilizando o terminal do Linux, após instalar o Python3.8:

1. Acesse o diretório em que o arquivo .py está salvo:
 - `$ cd user/local`
2. Instale as bibliotecas requeridas:
 - `$ pip3 install -r requirements.txt`

3. Execute o arquivo usando Python3.8

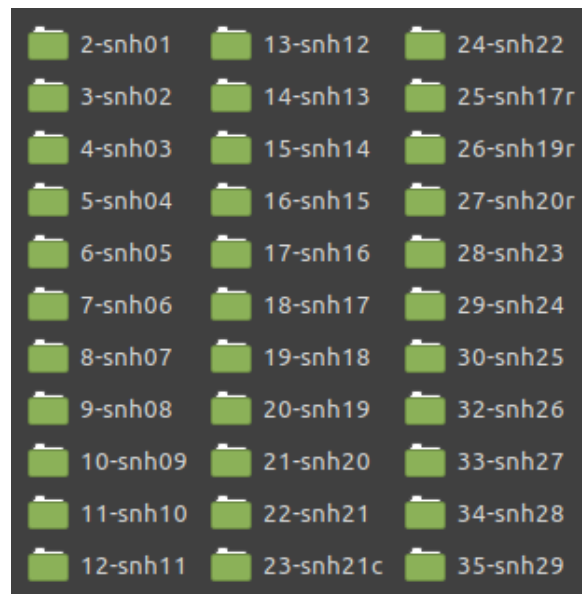
- `$ python3.8 script-anais-anpuh.py`

Bibliotecas e módulos

- **urllib.requests:** módulo do Python que ajuda a acessar urls. [Saiba mais.](#)
 - **os:** módulo do Python que permite manipular funções do sistema operacional. [Saiba mais.](#)
 - **bs4:** [Beautiful Soup](#) é uma biblioteca Python para extrair dados de arquivos HTML e XML.
 - **re:** [Regular Expressions](#) é um módulo do Python para operar com expressões regulares.
 - **pandas:** [Pandas](#) é uma biblioteca escrita em Python para manipulação e análise de dados.
-

Resultados

O script retorna para o usuário **todos os pdfs disponíveis em todas as páginas de todos os Simpósios Nacionais da Anpuh desde 1963 até 2017**. São criadas pastas com o número de cada evento para o armazenamento dos arquivos em PDF.



É importante notar que muitos papers não estão com pdf disponível no site, assim como nas edições mais antigas encontramos arquivos que contém vários papers num único PDF.

O script também gera um arquivo **CSV** (*comma-separated values*) contendo os seguintes valores para cada paper: **Autor(es)/Instituições, Título, Tipo, Evento, Ano, Link do Arquivo**. Esse arquivo pode ser aberto como uma planilha e trabalhado em banco de dados.

	Autor(es)/Instituições	Título	Tipo	Evento	Ano	Link do Arquivo
0	Maria Helena Capelato	os desafios da anpuh frente à crise brasileira: a luta pela preservação da demo	Conferência	XXIX Simpós	2017	https://anpuh.org.br/
1	Abner Neemias da Cruz (Univ	percursos políticos independentistas: a primeira geração diplomática, o impéri	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
2	A. Ricardo Abdalla (Pontificia	comensalidade e memória árabe na área central da cidade de são paulo	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
3	Acácia Regina Pereira (UERJ	currículo de história como responsável pela formação da identidade e memóri	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
4	Adalberto Paranhos (UNIVER	à flor da pele: pulsações do desejo feminino na música popular brasileira dos a	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
5	Adalberto Coutinho de Araújo	democracia e socialismo: propostas de políticas econômicas do operariado em	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
6	Adna Gomes Oliveira (UERJ)	o itamaraty, o corpo diplomático e a onu no início da guerra fria: a atuação bra	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
7	Adriana Aparecida Pinto (Uni	a imprensa e discurso sobre mulheres na imprensa rondonopolitana nos anos f	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
8	Adriana Maria de Souza Ziery	elementos religiosos da ascensão de d. joão i ao poder: o messias, o povo e a ci	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
9	Adriana Barreto de Souza (UI	jósé narcizo de magalhães e menezes, a profissão militar e as milícias de home	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
10	Aécio Lessa Macedo (Univers	as potencialidades do estudo da história local em sala de aula: análise de uma	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
11	Adriane Piovezan (FIES Facul	rituais fúnebres militares: o túmulo dos fuzileiros navais mortos na intentona i	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
12	Afonso Henrique Sant Ana B.	labaredas do rio - abordagens histórico-sociais do rio e do corpo de bombeiros	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
13	Aginaldo Kupper (Universida	futebol e contextos	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
14	Agenor Sarraf Pacheco (UNIV	a folia da cidade-floresta: patrimônio do afeto e lutas pela memória em melga	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
15	Agostinho Júnior Holanda Cr	a santa casa da misericórdia do maranhão e a intervenção dos presidentes-pro	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
16	Alan Christian de Souza Sant	lauro sodré em revista: textos, traços e retratos do senador paraense no perio	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
17	Airton José Cavenaghi (Anhe	mapas turísticos e sua representação cultural: marcel mauss e seu "ensaio sob	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
18	Aiman Jorge Henrique Franc	ivan ribeiro e a "via prussiana" no mundo rural brasileiro	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
19	Ailton José Morelli (Universi	a importância dos espaços nas memórias de infância	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
20	Alana Cavalcanti Cruz (UNIVE	memórias de moradores da rua duque de caxias, João Pessoa-PB, da primeira r	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
21	Alan Ricardo Duarte Pereira	trajetórias militares no império português: notas introdutórias sobre a família	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
22	Alberto Dias Mendes (UERJ)	jânio quadros e as influências de bandung	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
23	Alanna de Jesus Teixeira (UF	a representação do passado em anatóle france: literatura e história	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
24	Alec Ichiro Ito (USP)	a sucessão episcopal no bispado e diocese de congo e angola entre 1607 e 161	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
25	Aldina da Silva Melo (UNIVEF	a África no ensino de história no Brasil	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/
26	Alessandra Gonzalez de Carv	da "zona de contato" ao "limite antagônico": disputas entre criollos e indígenas	Comunicação	XXIX Simpós	2017	https://anpuh.org.br/

O script está funcionando perfeitamente. Qualquer alteração no site percebida pelos usuários ou sugestões de aprimoramento são bem vindas.

Licença

MIT licensed

Copyright (C) 2020 Eric Brasil, Gabriel Andrade, LABHD-UFBA