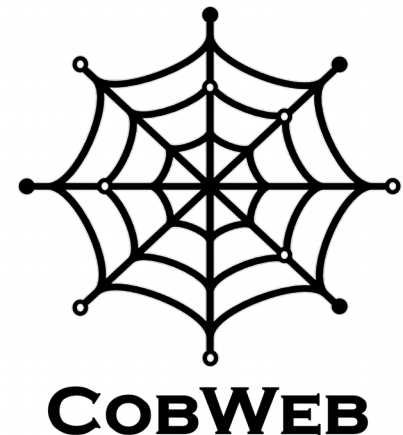


Sumário:

- Relatórios de Praga
- Definimos um logo para Start-Up
- Testes e Aprendizado de WebBot com Scrapy

Riscos:

- Tempo curto
- Falta de internet



WBS	ATIVIDADES	INICIO	FIM	FINALIZADO
1	Levantamento de contatos	3/4/2017	19/jun	19/jun
2	Definir funções	3/4/2017	19/jun	19/jun
3	Planejamento	3/4/2017	17/abr	17/abr
3.1	Realizar Project Charter	3/4/2017	3/4/2017	3/4/2017
3.2	Reunir fontes macro	4/4/2017	17/abr	17/abr
3.3	Reunião de ferramentas	4/4/2017	17/abr	17/abr
3.4	Definir curva de pertinencia	10/abr	17/abr	17/abr
4	Plataforma	18/abr	28/abr	28/abr
4.1	Definir plataformas	18/abr	21/abr	21/abr
4.2	Qualificação de fontes	24/abr	27/abr	02/abr
4.3	Reunião com grupo de desenvolvimento	28/abr	28/abr	21/abr
5	Desenvolvimento da Aplicação	1/mai	15/jun	
5.1	Manual	1/mai	29/mai	
5.1.1	Reunião de Ferramentas	1/mai	5/mai	
5.1.2	Preparação da Plataforma	2/mai	19/mai	
5.1.3	Alimentação do Banco de Dados	19/mai	24/mai	
5.1.4	Testes	24/mai	29/mai	
5.2	Automatização	29/mai	15/jun	
5.2.1	Gerar Plataforma de Armazenamento	29/mai	5/jun	
5.2.2	Gerar WebBot	29/mai	14/jun	
5.2.3	Testes	14/jun	15/jun	
6	Fechamento do Projeto	15/jun	19/jun	
6.1	Validação do Cliente	15/jun	16/jun	
6.2	Apresentação do Produto Final	16/jun	19/jun	


Scrapy | A Fast and Powerful Web Crawling Framework

← <https://scrapy.org>

Mais visitados Primeiros passos Veduca

Scrapy

Download Documentation Resources Community Companies FAQ Fork on Github




Scrapy

An open source and collaborative framework for extracting the data you need from websites. In a fast, simple, yet extensible way.

pypi v1.3.3 wheel yes coverage 84%

Build and run your web spiders

Install the latest version of Scrapy

 **Scrapy 1.3**

`$ pip install scrapy`

PyPI Conda Source

brito@brito: ~/Documentos/FACUL/PIP/Scrapy Programs/Teste 01

```
2017-05-08 10:01:40 [scrapy] INFO: Enabled item pipelines:
2017-05-08 10:01:40 [scrapy] INFO: Spider opened
2017-05-08 10:01:40 [scrapy] INFO: Crawled 0 pages (at 0 pages/min), scraped 0 items (at 0 items/min)
2017-05-08 10:01:40 [scrapy] DEBUG: Telnet console listening on 127.0.0.1:6023
2017-05-08 10:01:41 [scrapy] DEBUG: Crawled (200) <GET https://weather.com/> (referer: None)
2017-05-08 10:01:41 [CobWeb] DEBUG: Ola, sou CobWeb e visitei o site: https://weather.com/
2017-05-08 10:01:41 [scrapy] DEBUG: Scraped from <200 https://weather.com/> {'e seu tamanho': 414192, 'Sua url': 'https://weather.com/'}
2017-05-08 10:01:41 [CobWeb] DEBUG: Agora vou para: https://www.climatempo.com.br/
2017-05-08 10:01:42 [scrapy] DEBUG: Crawled (200) <GET https://www.climatempo.com.br/> (referer: https://weather.com/)
2017-05-08 10:01:42 [CobWeb] DEBUG: Visitei o climatempo via URL: https://www.climatempo.com.br/
2017-05-08 10:01:42 [scrapy] INFO: Closing spider (finished)
2017-05-08 10:01:42 [scrapy] INFO: Dumping Scrapy stats:
{'downloadloader/request_bytes': 450
```

(1) TecMundo - YouTube x

https://www.youtube.com/user/baixaki/videos

Mais visitados Primeiros passos Veduca

You Tube BR TechMundo

Início
Meu canal
Em alta
Inscrições 2

BIBLIOTECA

Histórico
Assistir mais tarde
Compras 7
Favorites

HOJE NO TECMUNDO

BRASIL LANÇA SATÉLITE DE COMUNICAÇÃO E DEFESA

Brasil lança satélite de comunicação e defesa - Hoje no TecMundo

35.423 visualizações • 2 dias atrás

```
youtube_spider.py x
1 import scrapy
2
3
4 def first(sel, xpath):
5     return sel.xpath(xpath).extract_first()
6
7
8 class YoutubeChannelLister(scrapy.Spider):
9     name = 'channel-lister'
10    youtube_channel = 'baixaki'
11    start_urls = ['https://www.youtube.com/user/%s/videos' % youtube_channel]
12
13    def parse(self, response):
14        for sel in response.css("ul#channels-browse-content-grid > li"):
15            yield {
16                'link': response.urljoin(first(sel, './h3/a/@href')),
17                'title': first(sel, './h3/a/text()'),
18                'views': first(sel, ".*ul/li[1]/text()"),
19            }
20
```

Line 10, Column 31

Tab Size: 4 Python

TecMundo.csv (~/.Documentos/FACUL/PIP/Scrapy Programs/Youtube 02) - gedit

Abrir Salvar

views	link	title
"35,410 views"	https://www.youtube.com/watch?v=c3Q8Wp0hA0A	Brasil lança satélite de comunicação e defesa - Hoje no TecMundo
"23,980 views"	https://www.youtube.com/watch?v=uJthCNGLXvM	Xiaomi Mi Note 2 vs OnePlus 3T - Comparativo [TecMundo]
"40,724 views"	https://www.youtube.com/watch?v=maL9fkeHEUK	Melhores apps para Android: (05/05/2017)
"23,597 views"	https://www.youtube.com/watch?v=oNtx2SdA5uk	Space X quer internet no espaço - Hoje no TecMundo
"33,448 views"	https://www.youtube.com/watch?v=aOLhn65no68	Samsung nos carros autônomos - Hoje no TecMundo
"15,592 views"	https://www.youtube.com/watch?v=ZcHHw1NL9ck	Quantum MUV Up: 5 prós e contras em relação à concorrência - TecMundo
"29,909 views"	https://www.youtube.com/watch?v=PlMyP9MNHwk	Galaxy J7 Prime: 5 prós e contras em relação à concorrência - TecMundo
"21,479 views"	https://www.youtube.com/watch?v=Mc-XWG_NfVo	Anac libera uso de drones no Brasil o registro é obrigatório - Drops
"74,330 views"	https://www.youtube.com/watch?v=yEG7apYsCXo	Microsoft revela o Windows 10 S - Hoje no TecMundo
"60,429 views"	https://www.youtube.com/watch?v=7TdUmknDH1E	Comparativo: Galaxy J7 Prime vs. Moto G5 Plus - TecMundo
"95,480 views"	https://www.youtube.com/watch?v=eu4ckZhcP6M	A história da Samsung - TecMundo
"43.501 views"	https://www.youtube.com/watch?v=DPTE0hGza1E	Comparativo: OnePlus 3T vs Samsung Galaxy S7 / S7 Edge - TecMundo

CSV Largura da tabulação: 8 Lin 12, Col 20 INS