



Coletor Intra-Site

Leonardo Alves (las3)
Leonardo Schettini (ljsa)
Marcos Barreto (msb5)

Domínio



Críticas de filme

1. Cineclick www.cineclick.com.br/
2. Cinemasim www.cinemasim.com.br
3. Cinema com rapadura www.cinecomrapadura.com.br/
4. Cinema em cena www.cinemaemcena.cartacapital.com.br/
5. El País www.brasil.elpais.com
6. Jornal da Paraíba www.jornaldaparaiba.com.br/
7. Omelete www.omelete.com.br
8. Plano aberto www.planoaberto.com.br/
9. Plano crítico www.planocritico.com/
10. Pocilga www.pocilga.com.br

Crawler



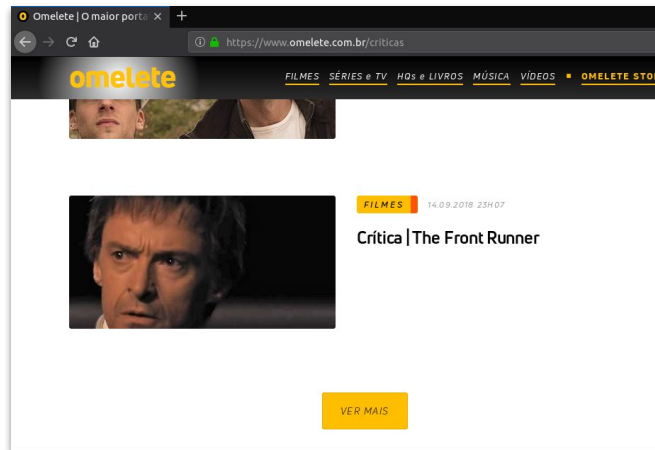
Ferramentas

- Python 3
- BeautifulSoup
- Requests



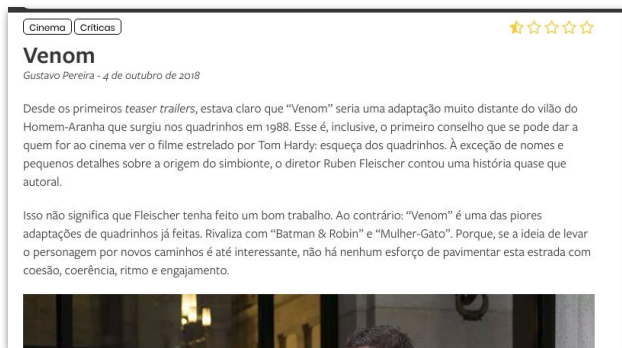
Página com javascript (Infinite Scrolling Pages)

- <https://www.omelete.com.br/criticas>
- Selenium



Links que não indicam bem o conteudo da pagina

- https://brasil.elpais.com/brasil/2018/07/26/cultura/1532607790_149183.html
- <http://www.planoaberto.com.br/2018/venom/>



Domínio	Qtd. Páginas relevantes	Qtd. Páginas visitadas	Harvest ratio
Cineclick	271	1000	0,271
Cinemasim	214	1000	0,214
Cinema Com Rapadura	353	1000	0,353
Cinema em Cena	131	1000	0,131
El País	78	1000	0,078
Jornal da Paraíba	103	1000	0,103
Omelete	90	1000	0,090
Plano Aberto	219	970	0,225
Plano Critico	637	1000	0,637
Pocilga	407	1000	0,407

Média geral: 0,329

Abordagem BFS - Resultados



Heurística 1

- Bag of words
 - **Good words:** crítica, resenha
 - **Index words:** “/page/”, “/index/”, “page=”, “/a/”
 - **Ok words:** cultura, filme, cinema
 - **Bad words:** notícia, série, lista, eventos, quadrinhos, temporada, álbum, literatura, televisão, música, podcast, sinopse
- Pesos
 - **Good words:** 3
 - **Index words:** 2
 - **Ok words:** 1
 - **Bad words:** -1
 - **Default:** 0

Domínio	Qtd. Páginas relevantes	Qtd. Páginas visitadas	Harvest ratio
Cineclick	833	1000	0,833
Cinemasim	190	372	0,510
Cinema Com Rapadura	412	583	0,706
Cinema em Cena	310	668	0,46
El País	0	37	0
Jornal da Paraíba	8	67	0,119
Omelete	763	1000	0,763
Plano Aberto	6	73	0,08
Plano Critico	559	1000	0,559
Pocilga	177	236	0,750

Média geral: 0,478

Heurística 1 - Resultados



Heurística 2

- Bag of words
 - Good words: crítica, resenha, “cultura/cinema/”.
 - Index words: “/page/”, “/index/”, “page=”, “/a/”.
 - Ok words: cultura, filme, cinema.
 - Bad words: notícia... governo, tag/discos/, share=, vida_urbana, colunas/, política, actualidad, cultura/agenda, tag/fecha/, economia, internacional, libros, prêmios, resenha-de-livro, clipe, autor/, emmy, tag/game_of_thrones, “[0-9]x[0-9]”.
- Pesos
 - Good words: 10
 - Index words: 3
 - Ok words: 1 | if pai == index word: 4
 - Bad words: -1
 - Default: 0 | if pai == index word: 2

Domínio	Qtd. Páginas relevantes	Qtd. Páginas visitadas	Harvest ratio
Cineclick	867	1000	0,867
Cinemasim	336	704	0,477
Cinema Com Rapadura	754	1000	0,754
Cinema em Cena	613	951	0,644
El País	24	122	0,196
Jornal da Paraíba	17	86	0,197
Omelete	760	1000	0,760
Plano Aberto	106	361	0,29
Plano Critico	533	1000	0,533
Pocilga	360	464	0,77

Média geral: 0,548

Heurística 2 - Resultados

Classificador



Roteiro

- Rotulação dos dados
- Balanceamento dos dados
- Pré-processamento
- Treinamento e teste
- Resultados
- Otimização



Rotulação dos dados

- No total 232 dados foram rotulados, dos quais 101 são documentos **relevantes** e 131 são **não relevantes**.
- Durante a leitura dos dados, um documento da classe **não relevante** apresentou erros e foi descartado.



Balanceamento dos dados

- Após os primeiros testes, os resultados apresentaram alta variância.
- Investigações apontaram para as instâncias escolhidas pelo *k-fold*.
- Para reduzir a variância um total de 101 instâncias de cada classe foi utilizado.



Pré-processamento

- Tokenizador com regex
 - e-mail, URLs, palavras coladas, telefone/cep, numerais, datas, palavras comuns e símbolos
- Filtro de tokens
- Stopwords (ignora siglas)
- Lower case
- Stemming



Treinamento e teste

- Feature selection: chi2 e mutual information.
- Para avaliação a técnica *k-fold* com $k=5$ foi utilizada.
- Adicionalmente, para normalizar a variância dos resultados, *5-fold* foi executado **30 vezes**. Salvo para o MLP, cujo tempo de treinamento impossibilitou tal estratégia.
- Para MLP, executamos o 5-fold **4 vezes** para cada tipo de input.

Naive Bayes

Aqui utilizamos o `MultinomialNB` pois ele aceita uma matriz esparça como input

```
In [35]: # Can train with sparse matrix
nb = MultinomialNB()

kfold_training(nb, term_document, labels, k=5, r=30)
```

```
Out[35]: {'precision': 0.6050000000000001,
          'recall': 0.6,
          'fscore': 0.6010025062656642,
          'accuracy': 0.6}
```

```
In [36]: # Can train with sparse matrix
nb = MultinomialNB()

kfold_training(nb, chi_term_document, labels, k=5, r=30)
```

```
Out[36]: {'precision': 0.8125,
          'recall': 0.5,
          'fscore': 0.47474747474747475,
          'accuracy': 0.5}
```

```
In [37]: # Can train with sparse matrix
nb = MultinomialNB()

kfold_training(nb, mutual_term_document, labels, k=5, r=30)
```

```
Out[37]: {'precision': 0.5611111111111111,
          'recall': 0.55,
          'fscore': 0.55,
          'accuracy': 0.55}
```

Resultados

Decision Tree

```
In [22]: dt = DecisionTreeClassifier()
```

```
kfold_training(dt, term_document, labels, k=5, r=30)
```

```
Out[22]: {'precision': 1.0, 'recall': 1.0, 'fscore': 1.0, 'accuracy': 1.0}
```

```
In [23]: dt = DecisionTreeClassifier()
```

```
kfold_training(dt, chi_term_document, labels, k=5, r=30)
```

```
Out[23]: {'precision': 0.906060606060606,  
          'recall': 0.9,  
          'fscore': 0.9007672634271099,  
          'accuracy': 0.9}
```

```
In [24]: dt = DecisionTreeClassifier(random_state=0)
```

```
kfold_training(dt, mutual_term_document, labels, k=5, r=30)
```

```
Out[24]: {'precision': 1.0, 'recall': 1.0, 'fscore': 1.0, 'accuracy': 1.0}
```

Resultados

SVM

```
In [38]: svcclassifier = SVC(C=0.5, kernel='rbf', degree=8, gamma=0.01, probability=True)
         kfold_training(svcclassifier, term_document, labels, k=5, r=30)
```

```
Out[38]: {'precision': 0.22562500000000002,
          'recall': 0.475,
          'fscore': 0.3059322033898305,
          'accuracy': 0.475}
```

```
In [39]: svcclassifier = SVC(C=0.5, kernel='rbf', degree=8, gamma=0.01, probability=True)
         kfold_training(svcclassifier, chi_term_document, labels, k=5, r=30)
```

```
Out[39]: {'precision': 0.22562500000000002,
          'recall': 0.475,
          'fscore': 0.3059322033898305,
          'accuracy': 0.475}
```

```
In [40]: svcclassifier = SVC(C=0.5, kernel='rbf', degree=8, gamma=0.01, probability=True)
         kfold_training(svcclassifier, mutual_term_document, labels, k=5, r=30)
```

```
Out[40]: {'precision': 0.18062499999999998,
          'recall': 0.425,
          'fscore': 0.2535087719298245,
          'accuracy': 0.425}
```

Resultados

Logistic regression

```
In [41]: # solvers: 'liblinear', 'sag', 'saga'
lg = LogisticRegression(random_state=0, solver='sag', multi_class='ovr')

kfold_training(lg, term_document, labels, k=5, r=30)
```

```
Out[41]: {'precision': 0.6528673835125448,
          'recall': 0.55,
          'fscore': 0.5171875,
          'accuracy': 0.55}
```

```
In [42]: # solvers: 'liblinear', 'sag', 'saga'
lg = LogisticRegression(random_state=0, solver='sag', multi_class='ovr')

kfold_training(lg, chi_term_document, labels, k=5, r=30)
```

```
Out[42]: {'precision': 0.7882352941176471,
          'recall': 0.55,
          'fscore': 0.496,
          'accuracy': 0.55}
```

```
In [43]: # solvers: 'liblinear', 'sag', 'saga'
lg = LogisticRegression(random_state=0, solver='sag', multi_class='ovr')

kfold_training(lg, mutual_term_document, labels, k=5, r=30)
```

```
Out[43]: {'precision': 0.6351648351648352,
          'recall': 0.6,
          'fscore': 0.591919191919192,
          'accuracy': 0.6}
```

Resultados

```
In [47]: mlp = MLPClassifier()  
         kfold_training(mlp, term_document, labels, k=5, r=4)
```

```
Out[47]: {'precision': 0.90625,  
          'recall': 0.875,  
          'fscore': 0.8769841269841271,  
          'accuracy': 0.875}
```

```
In [48]: mlp = MLPClassifier()  
         kfold_training(mlp, chi_term_document, labels, k=5, r=4)
```

```
Out[48]: {'precision': 0.9368421052631579,  
          'recall': 0.925,  
          'fscore': 0.9257142857142856,  
          'accuracy': 0.925}
```

```
In [49]: mlp = MLPClassifier()  
         kfold_training(mlp, mutual_term_document, labels, k=5, r=4)
```

```
Out[49]: {'precision': 0.8324808184143222,  
          'recall': 0.825,  
          'fscore': 0.824010056568196,  
          'accuracy': 0.825}
```

Resultados - MLP



Otimização

- Alta variância entre as execuções, mesmo rodando o 5-fold 30 vezes
- Tempo de treinamento

Random Forest

```
In [44]: rf = RandomForestClassifier(n_estimators=100, max_depth=10, random_state=200)

kfold_training(rf, term_document, labels, k=5, r=30)
```

```
Out[44]: {'precision': 0.934375,
          'recall': 0.925,
          'fscore': 0.9242857142857144,
          'accuracy': 0.925}
```

```
In [45]: rf = RandomForestClassifier(n_estimators=100, max_depth=10, random_state=200)

kfold_training(rf, chi_term_document, labels, k=5, r=30)
```

```
Out[45]: {'precision': 0.9235294117647058,
          'recall': 0.9,
          'fscore': 0.9026666666666667,
          'accuracy': 0.9}
```

```
In [46]: rf = RandomForestClassifier(n_estimators=100, max_depth=10, random_state=200)

kfold_training(rf, mutual_term_document, labels, k=5, r=30)
```

```
Out[46]: {'precision': 0.9258838383838384,
          'recall': 0.925,
          'fscore': 0.9248585795097423,
          'accuracy': 0.925}
```

Otimização

Extrator



Roteiro

- Extratores específicos
- Extrator geral
- Resultados
- Discussão



Extratores específicos

- Foram construídos utilizando o XPath dos elementos
 - Com o pacote lxml de python
- Em poucos casos, um único elemento continha toda a informação buscada
 - A maioria era separado, o nome do filme com os detalhes
- Cada extrator recebe a página HTML específica do site, e retorna um JSON com as informações extraídas

quanto com Owen Vaccaro, que faz Lewis.

O roteiro de Eric Kripke poderia explorar um pouco melhor o passado dos personagens, mas está mais interessado em introduzir os elementos da magia ou a dinâmica da casa de Jonathan. No entanto a história constrói de maneira satisfatória os participantes da narrativa, como ao mostrar as dificuldades do protagonista em conseguir se adaptar à nova vida em outra cidade, conflitos no colégio e outras coisas da vida comum de uma criança.

Em síntese, O Mistério do Relógio na Parede é uma grata surpresa na filmografia de Eli Roth e o diretor sai da sua zona de conforto explorando um outro lado como cineasta. Roth faz uma aventura de fantasia infanto-juvenil de qualidade, com personagens divertidos e um universo interessante que mistura magia e passado, com leves toques de nostalgia.

Uma frase: – Lewis: “Está usando um robe?”
– Jonathan: “É um quimono.”

Uma cena: A chegada de Lewis na casa do seu tio Jonathan.

Uma curiosidade: A casa onde a história se passa – presente no título original “The House with a Clock in Its Walls” – é inspirada em uma construção real: a Mansão Cronin, situada em Marshall, Michigan, cidade natal do autor do livro John Bellairs.



O Mistério do Relógio na Parede (The House with a Clock in Its Walls)

Direção: Eli Roth
Roteiro: Eric Kripke
Elenco: Jack Black, Cate Blanchett, Owen Vaccaro, Renée Elise Goldsberry, Sunny Suljic e Kyle MacLachlan
Gênero: Comédia, Fantasia, Família
Ano: 2018
Duração: 105 minutos



Ramon Prates

Analista de sistemas nascido em Salvador (BA) em 1980, mas atualmente morando em Brasília (DF). Cinema é sem dúvidas o meu hobby favorito. Assistio a filmes desde pequeno influenciado principalmente por meus pais e meu avô materno. Em seguida vem a música, principalmente rock e pop.

Compartilhe:

Pocilga

```
xpath_details = '//*[@starts-with(@id, "post-")]/div[2]/blockquote[2]/p'  
xpath_title = '//*[@starts-with(@id, "post-")]/div[2]/h4'
```

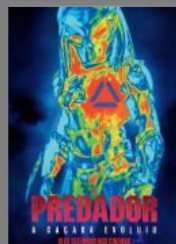
XPath

```
filmes = []  
for i in range(91,101):  
    filme = {'id': i}  
    tree = etree.parse(root + str(i) + '.html', parser)  
    title = tree.xpath(xpath_title)  
    title = get_element_text(title[0])  
    filme["nome"] = title.strip()  
    details = tree.xpath(xpath_details)  
    details = get_element_text(details[0]).split("\n")  
    for item in details:  
        if item.find(":") > -1:  
            filme[remover_acentos(item[item.find(":").lower()]] = item[item.find(":")+1:].strip()  
    filmes.append(filme)
```

filmes

```
[{'id': 91,  
  'nome': 'O Mistério do Relógio na Parede (The House with a Clock in Its Walls)',  
  'direcao': 'Eli Roth',  
  'roteiro': 'Eric Kripke',  
  'elenco': 'Jack Black, Cate Blanchett, Owen Vaccaro, Renée Elise Goldsberry, Sunny Suljic e Kyle MacLachlan',  
  'genero': 'Comédia, Fantasia, Família',  
  'ano': '2018',  
  'duracao': '105 minutos'},  
 ...]
```

Exemplo para o site pocilga.com.br



O PREDADOR

(THE PREDATOR)
2018, 101 MIN.

18



Gênero: Ação
Estreia: 13/09/2018

VOLTAR À PÁGINA INICIAL DO FILME

CineClick

```
titulo = '//*[@id="movie"]/div[2]/div/div/div[1]/div[2]/div[2]/h1'
ano_duracao='//*[@id="movie"]/div[2]/div/div/div[1]/div[2]/div[2]/p'
genero = '//*[@id="movie"]/div[2]/div/div/div[1]/div[2]/div[2]/div[2]/p[1]'
direcao='//*[@id="movie"]/div/ul'
details='//*[@id="movie"]/div/p'
elenco='//*[@id="movie"]/div/div/p'
```

ONDE ASSISTIR

FICHA TÉCNICA

NOTÍCIAS

CRÍTICAS

FOTOS E VÍDEOS

DIREÇÃO

Shane Black

EQUIPE TÉCNICA

Roteiro: Fred Dekker, Jim Thomas, John Thomas, Shane Black

Produção: Blondel Aidoo, John Davis, Lawrence Gordon

Fotografia: Larry Fong

Trilha Sonora: Henry Jackman

Estúdio: Canada Film Capital, Dark Castle Entertainment, Davis Entertainment, TSG Entertainment, Twentieth Century Fox Film Corporation

Montador: Harry B. Miller III

Distribuidora: Fox Film

ELENCO

Alfie Allen, Andrew Jenkins, Augusto Aguilera, Boyd Holbrook, Crystal Mudry, Edward James Olmos, Inka Malovic, Jacob Tremblay,

Um caso excepcional: cineclick.com.br



Extrator geral

- Navegação na DOM Tree em busca de um par candidato para a extração
 - Elenco e Direção são os pares que ao menos um está presente em todos os 10 sites
- Com esse par encontrado, é realizada uma busca “*bottom-up*” na tentativa de encontrar o elemento HTML “pai” dos demais pares para a extração
 - A partir do par candidato, é feita a busca em até 3 níveis da hierarquia, ou encontrar pelo menos:
 - Elenco e Diretor;
 - Elenco e Roteiro;
 - Diretor e Gênero;
 - Diretor e Roteiro.

Texto originalmente publicado como parte da cobertura do Plano Aberto para o 51º Festival de Brasília.
Para ler outros textos de nossa cobertura, clique [aqui](#).

Compartilhe isso:



Ficha Técnica



New Life S.A., 2018 – Brasil

Direção: André Carvalheira

Roteiro: André Carvalheira, Aurélio Aragão, Pablo Gonçalves

Elenco: Fernanda Rocha, Murilo Grossi, Catarina Accioly, Eric Naves, Edu Moares, Juliano Coacci, André Deca, Bianca Terraza, Wellington Abreu, Marcelo Pelucio

Fotografia: Krishna Schmidt

Design de Produção: Maira Carvalho



Wallace Andrioli

Historiador, pesquisador e crítico apaixonado por cinema, mais especificamente pelos filmes de Billy Wilder, Alfred Hitchcock, Clint Eastwood e Edward Yang.

▼<div class="details">

▶<p>...</p>

▼<p>

Direção:

" André Carvalheira"

</p>

▼<p>

Roteiro:

" André Carvalheira, Aurélio Aragão, Pablo Gonçalves"

</p>

▼<p>

Elenco:

" Fernanda Rocha, Murilo Grossi, Catarina Accioly, Eric Naves, Edu Abreu, Marcelo Pelucio"

</p>

▶<p>...</p>

▶<p>...</p>

</div>

Representação da estratégia



Extrator geral

- Para facilitar a detecção e extração de informação relevante, a estratégia utiliza uma lista de “top words”
 - Uma bag-of-words com os termos “key” dos pares a serem extraídos
 - *Diretor, Elenco, Gênero, Duração, Classificação, Roteiro...*
- De forma similar, uma lista de “bad words” fez-se necessário, para corrigir as “sujeiras” que vinha na extração
 - *Ficha técnica, Equipe técnica...*



Extrator geral

- Os sites têm formas distintas de informar o título do filme
 - Alguns dispõe juntamente com as informações técnicas
 - Outros apenas no título da página/artigo
- Assim, a estratégia checa se o título veio na extração tradicional (elemento HTML pai)
 - Caso sem sucesso, é checado 3 condições na DOM Tree. Os elementos que:
 - O atributo *class* é “title”
 - O atributo *class* que termina com “titulo”
 - O atributo *class* que termina com “title”
 - Infelizmente dois sites não disponibilizavam de forma identificável o nome do filme
 - Separado dos detalhes técnicos, ou sem um atributo identificador: *id/class*

N = Total de extrações possíveis
 E = Total de pares extraídos pelo sistema
 C = Total de pares extraídos
 corretamente

Domínio	N	E	C	Recall	Precision	F1
Cineclick	129	99	89	0.689	0.898	0.780
Cinemasim	69	69	65	0.942	0.942	0.942
Cinema Com Rapadura	50	30	20	0.4	0.666	0.5
Cinema em Cena	113	94	84	0.743	0.893	0.811
El País	49	49	47	0.959	0.959	0.959
Jornal da Paraíba	58	58	58	1	1	1
Omelete	60	38	37	0.616	0.973	0.755
Plano Aberto	58	56	56	0.965	1	0.982
Plano Critico	50	50	50	1	1	1
Pocilga	70	70	65	0.928	0.928	0.928
Total	706	613	571	0.808	0.931	0.865

Extrator geral - Resultados



Análise do resultado - El País

- O extrator conseguiu pegar todos os pares possíveis
- Porém em duas páginas, o nome do filme não estava no mesmo elemento HTML que os detalhes técnicos
 - A heurística descrita para capturar o nome do filme utilizando o atributo *class* não teve sucesso
- Resultado da informação nome extraída incorretamente
 - **Formato: O que foi extraído != O que deveria ser**
 - Seleccione Edição != 'Pantera Negra'
 - Seleccione Edição != TULLY

Domínio	N	E	C	Recall	Precision	F1
El País	49	49	47	0.959	0.959	0.959



Análise do resultado - Omelete

- Os detalhes técnicos são divididos em dois elementos HTML, e em apenas um contém as informações de Direção, Elenco e Roteiro, por exemplo
 - Essas são as consideradas importantes pela estratégia
 - Assim, o outro elemento HTML não é recuperado
- Em uma página o nome do filme havia o ano de lançamento, o que diferiu na comparação
 - **Formato: O que foi extraído != O que deveria ser**
 - O Predador != O Predador (2018)

Domínio	N	E	C	Recall	Precision	F1
Omelete	60	38	37	0.616	0.973	0.755

O resultado é um suspense em que os truques da interface online se prestam mais à exposição do que a uma imersão narrativa em si (de novo, *Amizade Desfeita* é mais forte nesse sentido), mas graças a sacadas pontuais e uma reviravolta final realmente engenhosa, Chaganty faz um whodunit que entrega o que promete.

Buscando...

Searching



ANO: 2018

DIREÇÃO: Aneesh Chaganty

ROTEIRO: Sev Ohanian, Aneesh Chaganty

ELENCO: Sara Sohn, John Cho

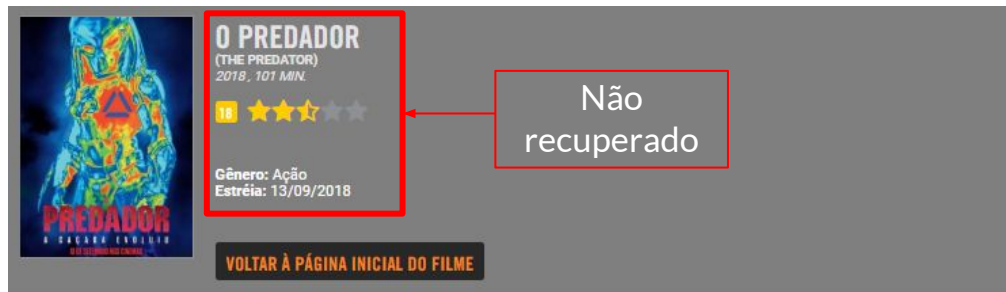
Análise do resultado - Omelete



Análise do resultado - Cineclick

- As páginas têm duas partes de detalhamento do filme, a estratégia recupera apenas um
 - A ficha técnica contém as informações consideradas importantes
 - Duração, ano e gênero estão na outra parte
- O nome do filme não está no mesmo elemento HTML dos detalhes técnicos, e a heurística recupera apenas alguns, mas na comparação é case sensitive
 - **Formato: O que foi extraído != O que deveria ser**
 - O Predador != O PREDADOR

Domínio	N	E	C	Recall	Precision	F1
Cineclick	129	99	89	0.689	0.898	0.780



ONDE ASSISTIR	FICHA TÉCNICA	NOTÍCIAS	CRÍTICAS	FOTOS E VÍDEOS
---------------	---------------	----------	----------	----------------

DIREÇÃO
Shane Black

EQUIPE TÉCNICA
Roteiro: Fred Dekker, Jim Thomas, John Thomas, Shane Black
Produção: Blondel Aidoo, John Davis, Lawrence Gordon
Fotografia: Larry Fong
Trilha Sonora: Henry Jackman
Estúdio: Canada Film Capital, Dark Castle Entertainment, Davis Entertainment, TSG Entertainment, Twentieth Century Fox Film Corporation
Montador: Harry B. Miller III
Distribuidora: Fox Film

ELENCO
Alfie Allen, Andrew Jenkins, Augusto Aguilera, Boyd Holbrook, Crystal Mudry, Edward James Olmos, Inka Malovic, Jacob Tremblay,

Recuperado

Análise do resultado - Cineclick



Análise do resultado - Cinema em cena

- Caso similar ao site “Cineclick”, existem duas áreas com detalhamento do filme, a estratégia recupera apenas um
 - Ano e distribuidora não são extraídos
- Esse é um dos dois sites que não disponibiliza de forma identificável o nome do filme
 - Assim, para todas as páginas testadas (10), nenhum foi extraído o nome

Domínio	N	E	C	Recall	Precision	F1
Cinema em Cena	113	94	84	0.743	0.893	0.811

Não
recuperado

CINEMA EM CENA			Criticas	Cinema ▾	Notícias	Casts ▾	Blog
Datas de Estreia:		Nota:					
Brasil	Exterior	Crítico					
09/08/2018	06/04/2018	★★★★☆					
Distribuidora							
Supo Mungan Films							

Recuperado

Direção
Lynne Ramsay
Elenco
Joaquin Phoenix , Ekaterina Samsonov , Alex Manette , John Doman , Dante Pereira-Olson , Judith Roberts , Alessandro Nivola
Roteiro
Lynne Ramsay
Produção
Pascal Caucheteux , Rebecca O'Brien , Lynne Ramsay , Rosa Attab , James Wilson
Fotografia
Thomas Townend
Música
Jonny Greenwood

A cineasta Lynne Ramsay, que diretores bem nesses é sua capacidade de criar simbolismos

Vejamos, por exemplo, absolutamente brilhante expectativas no público fugindo de padrões: experimentar frustra-se logo percebemos a constatar como os figurinos elegantes, moletins e casacos longos, constrói um

Ah, sim: e mesmo te

O fato é que, ao longo das passagens a sala de cinema, ao contrário responde Joe depois visto como virtude e

São hesitações com frequência trazem atormentadas que por exemplo). Exibir como um indivíduo com personalidade; um : condicionou a seguir com a mãe idosa re: com ele um passado as próprias causas.

Enquanto isso, Lynn longa anterior, *Preciosa* fotografia – e aqui e

Análise do resultado - Cinema em cena




Análise do resultado - Cinema com rapadura

- Esse é um caso específico porque as informações de Ano e Diretor estão implícitas na região da extração de informação
- O extrator específico para esse site foi modelado para tratar esses dois casos
- O nome do filme é recuperado pelo extrator geral, porém difere do extrator específico por conta o ano
 - **Formato: O que foi extraído != O que deveria ser**
 - O Predador (2018) != O Predador

Domínio	N	E	C	Recall	Precision	F1
Cinema Com Rapadura	50	30	20	0.4	0.666	0.5

Neste processo, “Buscando...” se encontra como um suspense que, embora experimental, consegue apelar para o grande público através de seu didatismo. Tentando surpreender narrativa e visualmente, tem mais sucessos do que fracassos, e leva seu público a refletir sobre seus espelhos negros e como deixamos toda uma vida – entre frustrações, relações humanas e fugas – dentro destes aparelhos.



BUSCANDO... (2018)
Searching - Aneesh Chaganty

Após uma jovem de 16 anos desaparecer, seu pai David Kim (John Cho) pede ajuda às autoridades locais. Sem sucesso, após 37 horas, David decide invadir o computador de sua filha para procurar pistas que possam levar ao seu paradeiro.

Roteiro: Aneesh Chaganty, Sev Ohanian
Elenco: John Cho, Debra Messing, Michelle La, Kya Dawn Lau, Megan Liu, Alex Jayne Go, Sara Sohn, Joseph Lee, Steven Michael Eich, Ric Sarabia, Sean O'Bryan, Thomas Barbusca, Roy Abramsohn, Brad Abrell, Courtney Lauren Cummings, Benjamin J. Cain Jr., Colin Woodell

com Rapadura

Entrar

Compartilhar

Ordenar por Mais recentes

Análise do resultado - Cinema com rapadura



Análise do resultado - Plano aberto

- Dentre as páginas testadas, houve apenas um caso específico onde o filme tinha apenas Diretor como uma das informações importantes
 - Como explicado anteriormente, para identificar o elemento HTML com os detalhes técnicos do filme, é considerado pelo menos **duas** informações
- Para as demais páginas, o extrator tem um ótimo recall e precision

Domínio	N	E	C	Recall	Precision	F1
Plano Aberto	58	56	56	0.965	1	0.982

Compartilhe isso:



Ficha Técnica



Excelentísimos, 2018 - Brasil

Direção: Douglas Duarte



Wallace Andrioli

Historiador, pesquisador e crítico apaixonado por cinema, mais especificamente pelos filmes de Billy Wilder, Alfred Hitchcock, Clint Eastwood e Edward Yang.

Análise do resultado - Plano aberto



Análise do resultado - Cinema sim

- O único problema é no momento de extrair o nome do filme de algumas páginas
 - **Formato:** O que foi extraído != O que deveria ser
 - As Herdeiras | CRÍTICA != As Herdeiras

Domínio	N	E	C	Recall	Precision	F1
Cinemasim	69	69	65	0.942	0.942	0.942



Análise do resultado - Pocilga

- O único problema é no momento de extrair o nome do filme de algumas páginas
 - **Formato:** O que foi extraído != O que deveria ser
 - A Primeira Noite de Crime != A Primeira Noite de Crime (The First Purge)

Domínio	N	E	C	Recall	Precision	F1
Pocilga	70	70	65	0.928	0.928	0.928

Obrigado!



Github do Projeto