



Busca por livros

Cinthyra (cml2)
Eládia Cristina Corrêa (ecmac)
Gabriela (glm)



CRAWLER

Sites coletados:

- Amazon
- Livraria Saraiva (15 sites)
- Livraria da Travessa (261 sites)
- Livraria Traça (102 sites)
- Cia. dos Livros (22 sites)
- Livraria Cultura (10 sites)
- Disal (45 sites)
- Extra
- Submarino
- Americanas

Estratégia usada: busca em largura



Saraiva

```
#gets links from the current page
def get_links(self, url):
    html = self.get_html(url)
    parsed = urlparse(url)
    base = f"{parsed.scheme}://{parsed.netloc}"
    links = re.findall('(<a\s+(?:[>]*?\s+)?href="([^\"]*[\d]+)/p+)', html)
    for i, link in enumerate(links):
        if not urlparse(link).netloc:
            link_with_base = base + "/" + link
            links[i] = link_with_base
    return set(filter(lambda x: 'mailto' not in x, links))
```

Harvest ratio (busca em largura): 12/16 = 75%



Travessa

```
#gets links from the current page
def get_links(self, url):
    html = self.get_html(url)
    parsed = urlparse(url)
    base = f"{parsed.scheme}://{parsed.netloc}"
    links = re.findall('\"<a\s+(?:[^\"])*?\s+href=\"([^\"]*)/artigo/+([^\"]*)\"', html)
    for i, link in enumerate(links):
        if not urlparse(link).netloc:
            link_with_base = base + "/" + link
            links[i] = link_with_base
    return set(filter(lambda x: 'mailto' not in x, links))
```

Harvest ratio (busca em largura) 261/261 = 100%

Traça

```
#gets links from the current page
def get_links(self, url):
    html = self.get_html(url)
    parsed = urlparse(url)
    base = f"{parsed.scheme}://{parsed.netloc}"
    links = re.findall('(<a\s+(?:[>]*?\s+)?href="([<"]*/livro/+[<"]*)"', html)
    for i, link in enumerate(links):
        if not urlparse(link).netloc:
            link_with_base = base + "/" + link
            links[i] = link_with_base
    return set(filter(lambda x: 'mailto' not in x, links))
```

Harvest ratio (busca em largura): $102/102 = 100\%$

Cia. dos Livros

```
#gets links from the current page
def get_links(self, url):
    html = self.get_html(url)
    parsed = urlparse(url)
    base = f"{parsed.scheme}://{parsed.netloc}"
    links = re.findall('(<a\s+(?:[>]*?\s+)?href="([^\"]*/produto/+\w-)*[\d]+[/"]*)"', html)
    for i, link in enumerate(links):
        if not urlparse(link).netloc:
            link_with_base = base + "/" + link
            links[i] = link_with_base
    return set(filter(lambda x: 'mailto' not in x, links))
```

Harvest ratio (busca em largura): $22/22 = 100\%$



Livraria Cultura

```
#gets links from the current page
def get_links(self, url):
    html = self.get_html(url)
    parsed = urlparse(url)
    base = f"{parsed.scheme}://{parsed.netloc}"
    links = re.findall('(<a\s+(?:[^\>]*?\s+)?href="([^\"]*p/livros/[\w/]+[\w-]+[\w\d]+[^\s/"]*)"' , html)
    for i, link in enumerate(links):
        if not urlparse(link).netloc:
            link_with_base = base + "/" + link
            links[i] = link_with_base
    return set(filter(lambda x: 'mailto' not in x, links))
```

Harvest ratio (busca em largura): $9/10 = 90\%$

Disal

```
#gets links from the current page
def get_links(self, url):
    html = self.get_html(url)
    parsed = urlparse(url)
    base = f"{parsed.scheme}://{parsed.netloc}"
    links = re.findall('(<a\s+(?:[>]*?\s+)?href="([^\"]*produto/+[\d]+[\w]*[/"]*)"', html)
    for i, link in enumerate(links):
        if not urlparse(link).netloc:
            link_with_base = base + "/" + link
            links[i] = link_with_base
    return set(filter(lambda x: 'mailto' not in x, links))
```

Harvest ratio (busca em largura): 45/45 = 100%



Classificação

- Código em Python
- BeautifulSoup para extrair tokens do HTML
- scikit-learn e Weka para classificar
- 244 páginas “positivas” e 244 páginas “negativas”
 - Amazon
 - Cia dos Livros
 - Disal
 - Estante Virtual
 - Extra
 - Livraria Cultura
 - Saraiva
 - Shoptime
 - Livraria Traça
 - Livraria da Travessa



CLASSIFICAÇÃO

- Bag of words
- Feature selection por frequência (pelo menos 50)
- Stemming (singular + plural)
- Remoção de números e tokens de tamanho pequeno (4+)
- Remoção de stopwords
- Matriz com frequências por documento e somente 0 ou 1
- Weka (5-fold cross validation)
 - MLP
 - SMO Reg
- scikit-learn (train 70%, test 30%)
 - Gaussian Naïve Bayes
 - Decision Tree
 - Logistic Regression (com 5-fold cross validation)



Gaussian Naïve Bayes (frequências)



```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(X_train, y_train)
classification = nb.predict(X_test)
score = nb.score(X_test, y_test)
print(score)
```



```
0.8163265306122449
```



Gaussian Naïve Bayes (0 ou 1)

```
[134] nb2 = GaussianNB()  
      nb2.fit(X_train2, y_train2)  
      classification2 = nb2.predict(X_test2)  
      score2 = nb2.score(X_test2, y_test2)  
      print(score2)
```

```
0.7891156462585034
```



Decision Tree (frequências)

```
▶ from sklearn import tree
dt = tree.DecisionTreeClassifier()
dt.fit(X_train, y_train)
classfDT = dt.predict(X_test)
scoreDT = dt.score(X_test, y_test)

print(scoreDT)
```

```
📄 0.9455782312925171
```



Decision Tree (0 ou 1)

```
[135] dt2 = tree.DecisionTreeClassifier()  
      dt2.fit(X_train2, y_train2)  
      classfDT2 = dt2.predict(X_test2)  
      scoreDT2 = dt2.score(X_test2, y_test2)  
  
      print(scoreDT2)
```

```
0.9455782312925171
```



Logistic Regression (frequências)

com 5-fold cross validation

```
[123] from sklearn.linear_model import LogisticRegressionCV
lr = LogisticRegressionCV(solver='sag', cv=5, max_iter=4000)
lr.fit(X_train, y_train)
classfLR = lr.predict(X_test)
scoreLR = lr.score(X_test, y_test)

print(scoreLR)
```

```
0.9387755102040817
```



Logistic Regression (0 ou 1)

com 5-fold cross validation



```
lr2 = LogisticRegressionCV(solver='sag', cv=5, max_iter=4000)
lr2.fit(X_train2, y_train2)
classfLR2 = lr2.predict(X_test2)
scoreLR2 = lr2.score(X_test2, y_test2)

print(scoreLR2)
```



```
0.9795918367346939
```


Multilayer Perceptron (frequências)

com attribute selection

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.898
Mean absolute error	0.1183
Root mean squared error	0.2206
Relative absolute error	23.6223 %
Root relative squared error	44.0495 %
Total Number of Instances	488

No.		
1	<input checked="" type="checkbox"/>	preto
2	<input checked="" type="checkbox"/>	jogo
3	<input checked="" type="checkbox"/>	bege
4	<input checked="" type="checkbox"/>	vermelha
5	<input checked="" type="checkbox"/>	mistério
6	<input checked="" type="checkbox"/>	rosa
7	<input checked="" type="checkbox"/>	isbn
8	<input checked="" type="checkbox"/>	milagre
9	<input checked="" type="checkbox"/>	asin
10	<input checked="" type="checkbox"/>	disco
11	<input checked="" type="checkbox"/>	encadernação
12	<input checked="" type="checkbox"/>	fabricante
13	<input checked="" type="checkbox"/>	livro
14	<input checked="" type="checkbox"/>	design
15	<input checked="" type="checkbox"/>	cabo
16	<input checked="" type="checkbox"/>	vídeo
17	<input checked="" type="checkbox"/>	CLASS

Multilayer Perceptron (0 ou 1)

com attribute selection

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.9264
Mean absolute error	0.0705
Root mean squared error	0.1913
Relative absolute error	14.0888 %
Root relative squared error	38.2023 %
Total Number of Instances	488

No.	
1	<input type="checkbox"/> ilustração
2	<input type="checkbox"/> assunto
3	<input type="checkbox"/> português
4	<input type="checkbox"/> isbn
5	<input type="checkbox"/> comum
6	<input type="checkbox"/> prateleira
7	<input type="checkbox"/> disney
8	<input type="checkbox"/> crash
9	<input type="checkbox"/> garrafa
10	<input type="checkbox"/> full
11	<input type="checkbox"/> editora
12	<input type="checkbox"/> volume
13	<input type="checkbox"/> brochura
14	<input type="checkbox"/> pacote
15	<input type="checkbox"/> size
16	<input type="checkbox"/> preto
17	<input type="checkbox"/> vida
18	<input type="checkbox"/> sistema
19	<input type="checkbox"/> milagre
20	<input type="checkbox"/> morte
21	<input type="checkbox"/> CLASS



SMO reg (frequências)

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.892
Mean absolute error	0.1466
Root mean squared error	0.2292
Relative absolute error	29.2741 %
Root relative squared error	45.7711 %
Total Number of Instances	488

```
SMOreg -C 1.0 -N 0 -I "weka.classifiers.functions.supportVector.RegSMOImproved -L 0.001 -W 1 -P 1.0E-12 -T 0.001 -V" -K "weka.classifiers.functions.supportVector.NormalizedPolyKernel -C 250007 -E 2.0
```



SMO reg (0 ou 1)

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.9443
Mean absolute error	0.106
Root mean squared error	0.1678
Relative absolute error	21.169 %
Root relative squared error	33.515 %
Total Number of Instances	488

```
SMOreg -C 1.0 -N 0 -I "weka.classifiers.functions.supportVector.RegSMOImproved -L 0.001 -W 1 -P 1.0E-12 -T 0.001 -V" -K "weka.classifiers.functions.supportVector.NormalizedPolyKernel -C 250007 -E 2.0
```



Classificação

	GNB	DT	LR	MLP	SMO
frequências	0.816	0.945	0.938	0.898	0.892
0 ou 1	0.789	0.945	0.975	0.926	0.944



Classificação - accuracy, precision, recall

	GNB	DT	LR	MLP	SMO
frequências	0.816	0.945	0.938		
0 ou 1	0.789	0.945	0.979		

	GNB	DT	LR	MLP	SMO
frequências	0.775	0.935	0.913		
0 ou 1	0.741	0.925	0.986		

	GNB	DT	LR	MLP	SMO
frequências	0.907	0.9605	0.973		
0 ou 1	0.907	0.973	0.973		



Classificador - problemas

- Tratar caracteres especiais e codificação
- Escolher como fazer o stemming
- Pouca experiência com Python



WRAPPER

O Wrapper coleta todos as informações de um nó pai, insere em uma lista

São eliminados caracteres da lista de modo que só reste o texto

As informações são coletadas da lista

Elementos fora do nó pai, são coletados através das tags específicas

WRAPPER SUBMARINO

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	1808: Como uma rainha	Laurentino Gomes	384	Globo Livros	(Português)	Capa Comum	978-8525057518
1	Box - As Melhores Aventuras Do Sítio Do Picapa...	Monteiro Lobato	640	Nova Fronteira	(Português)	Capa dura	978-8520943519
2	Cotoco. O Diário De Um Garoto De 13 Anos	John Van de Ruit	390	Intrinseca	(Português)	Capa Comum	978-8598078854
3	Doctor Who: 12 Doutores, 12 histórias	Alex Scarrow	480	Fantástica Rocco	(Português)	Capa Comum	978-8568263044
4	Freud (1900) A interpretação dos sonhos: Obras...	Sigmund Freud	736	Companhia das Letras	(Português)	Capa dura	978-8535932218
5	Good Omens: Belas Maldições	Neil Gaiman	364	Bertrand Brasil	(Português)	Capa Comum	978-8528624021
6	Os Mitos de Cthulhu - Volume Único (Exclusivo ...	Esteban Maroto	92	Pipoca e Nanquim (7 de junho de 2019)	(Português)	Capa dura	978-8593695261
7	Redes de Computadores e a Internet: Uma Aborda...	Jim Kurose	656	Pearson Universidades	(Português)	Capa Comum	978-8581436777
8	Sobotta - Atlas de Anatomia Humana - 3 Volumes	Friedrich Paulsen	1168	Guanabara Koogan	(Português)	Capa dura	978-8527732376

Recall: 1.0
Precision: 1.0
F_Measure = 1.0
C = 9
N = 9
E = 9

WRAPPER SARAIVA

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	Alice No País Das Maravilhas (Classic Edition)	Carroll,Lewis	208		Português	Capa dura	9788594541758
1	Steve Jobs - A Biografia	Isaacson,Walter, Isaacson,Walter	624		Português	Brochura	9788535919714
2	Vade Mecum Compacto - Brochura - 21ª Ed. 2019	Editora Saraiva	2127		Português	Brochura	9788553603244
3	It - A Coisa	King,Stephen	1104		Português	Brochura	9788560280940
4	Como Fazer Amigos e Influenciar Pessoas		264		Português	Brochura	9788504018028
5	O Guia do Mochileiro Das Galáxias - Série o Mo...	Adams,Douglas	208		Português	Brochura	9788599296578
6	O Livrinho do Bebê - Animais, Palavras, Contrá...	Books,Yoyo	128		Português	Brochura	9789463041706
7	Star Wars - O Arquivo Rebelde	Wallace,Daniel	176		Português	Brochura	9788528622973

```
Recall: 0.0
Precision: 0.0
F_Measure = 0
C = 0
N = 8
E = 8
```

WRAPPER

LIVRARIA DA TRAVESSA

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	A COR DA MAGIA	Terry Pratchett	232	Conrad	Português	Brochura	9788587193391
1	CYNDI MINHA HISTORIA	Cyndi Lauper Jancee Dunn Aline Naomi Sas...	352	Belas-Letras	Português	Brochura	9788581744988
2	D. PEDRO II A HISTORIA NAO CONTADA - O ULTIMO ...	Paulo Rezzutti	576	Leya	Português	Brochura	9788577346776
3	JAPONES PARA LEIGOS	Eriko Sato Samantha Batista	416	Alta Books	Português	Brochura	9788550802336
4	LONELY PLANET ARGENTINA (INCLUI URUGUAI)		640	Globo Livros	Português	Brochura	9788525068125
5	MADLINE FINN E OS CÃES DO ABRIGO	Lisa Papp Lenice Bueno	32	Salamandra	Português	Brochura	9788516117979
6	O PODER DA AUTORRESPONSABILIDADE A FERRAMENTA ...	Paulo Vieira	160	Gente	Português	Brochura	9788545202219
7	CRONICAS PARA LER EM QUALQUER LUGAR	Gregorio Duvivier Maria Ribeiro Xico Sa	112	Todavia	Português	Brochura	9786580309375
8	SO PARA UM ALIMENTAÇÃO SAUDAVEL PARA QUEM MORA...	Rita Lobo	168	Senac SP	Português	Capa dura	9788539629510
9	VIDAS SECAS - 140ªED.(2019)	Graciliano Ramos	176	Record	Português	Brochura	9788501114785

Recall: 0.9
Precision: 0.9
F_Measure = 0.9
C = 9
N = 10
E = 10

WRAPPER TRAÇA LIVRARIA

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	A Casa Dos Budas Ditosos	João Ubaldo Ribeiro	164	Objetiva	Português	Brochura	9788573022391
1	Cavaleiros Do Zodíaco Vol 1	Masami Kurumada	110	Ano	Português	Brochura	
2	Cem Anos De Solidão	Gabriel Garcia Marquez	368	1967	Português	Brochura	
3	Crianças Famosas: Schubert	Ann Rachlin E Susan Hellard	24	Editora	Português	Grampeado	9788574164632
4	Indiozinhos	Mônica Haibara	12	FTD	Português	Grampeado	
5	Livro Da Esperança	Francisco Cândido Xavier - Emmanuel	240	1984	Português	Brochura	
6	Mozart Don Giovanni - A Vida De Mozart	Lorenzo Da Ponte - Henry Stendhal	200	1991	Português	Brochura	
7	O Colapso Do Universo	Isaac Asimov	210	1987	em(Brasil)	Capa Dura	
8	Políticas Territoriais Na Amazônia	Neli Aparecida De Mello	412	2006	Português	Brochura	9798574196205
9	Relações Exteriores Do Brasil Contemporâneo	Danielly Silva Ramos Becard	224	Vozes	Português	Brochura	9788532638861

Recall: 0.4
Precision: 0.4
F_Measure = 0.4
C = 4
N = 10
E = 10

WRAPPER CIA DOS LIVROS

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	Afirmar-se com Nietzsche	Balthasar Thomass	208	Vozes Nobilis	Português	Brochura	9788532660480
1	Audácia Dessa Mulher, A	Ana Maria Machado	408	Alfaguara	Português	Brochura	9788556520876
2	Avaliação: Mito e Desafio	Jussara Hoffmann	158	MEDIACAO	Português	Brochura	9788577061273
3	Breve História do Feminismo no Contexto Euro-a...	Antje Schrupp	88	EDGARD BLUCHER	Português	Brochura	9788521214038
4	Corpo Fala, O: A Linguagem Silenciosa da Comun...	Pierre Weil	287	VOZES	Português	Brochura	9788532602084
5	Cura do Ciúme, A: Aprenda a Confiar, Supere a ...	Robert L. Leahy	192	Artmed	Português	Brochura	9788582715550
6	Dom Casmurro - Coleção Grandes Mestres da Lite...	Machado De Assis	192	LAROUSSE	Português	Brochura	9788581863009
7	Em Busca do Texto Perfeito: Questões Contempor...	Ana Elisa Ribeiro	96	GULLIVER	Português	Brochura	9788565432436
8	Manual do Arquiteto Descalço - Capa Dura	Johan Van Lengen	720	B4 Editores	Português	Capa dura	9788565358668
9	Feminismo É Feminino?, O	Maíra Marcondes Moreira	204	Annablume	Português	Brochura	9788539109456
10	Gramática de Espanhol Para Brasileiros	Esther Maria Milani	432	Saraiva - Didaticos	Português	Brochura	9788502132481
11	Nem Só De Ciência Se Faz A Cura: O Que Os Paci...	Protasio L. Da Luz	304	Manole	Português	Brochura	9788578683788

Recall: 1.0
Precision: 1.0
F_Measure = 1.0
C = 12
N = 12
E = 12

WRAPPER EXTRA

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	Aromaterapia e as Emoções: Como Usar Óleos Ess...	Shirley Price	322	Bertrand	Português	Brochura	9788528608519
1	Livro - Once Upon A Time - Irmãos Grimm	Irmãos Grimm	272	Planeta	Português	Brochura	9788542203080
2	Livro - Os Miseráveis - Walcyr Carrasco	Walcyr Carrasco; Victor Hugo	214	Moderna	Português	Brochura	8516079732
3	Livro - Panelinha - Receitas Que Funcionam - R...	Rita Lobo	400	Senac	Português	Brochura	8539602776
4	Livro - A Pequena Filosofia da Mafalda - Vamos...	Joaquim Salvador Lavado (Quino)	36	Martins	Português	Brochura	9788580632156
5	Livro - A Seleção: 35 Garotas e uma Coroa - Vo...	Kiera Cass	368	Seguinte	Português	Brochura	9788565765015
6	Livro - Como Eu Era Antes de Você - Jojo Moyes	Jojo Moyes	320	Intrínseca	Português	Brochura	8580573297
7	Livro - Elas Podem... e Devem: o Livro que vai...	Flaviane Brandemberg	114	Ser	Português	Brochura	8563178393
8	Livro - Em Busca de Sentido: Um Psicólogo no C...	Viktor E. Frankl	186	Vozes	Português	Brochura	8532606261
9	Livro - Eu, Fernando Pessoa: em Quadrinhos - S...	Susana Ventura	72	Peirópolis	Português	Brochura	9788575963050

```
Recall: 1.0
Precision: 1.0
F_Measure = 1.0
C = 10
N = 10
E = 10
```

WRAPPER CULTURA

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	MAIS ESPERTO QUE O DIABO	HILL, NAPOLEON	200	CDG	PORTUGUÊS	BROCHURA	8568014003
1	O CAIBALION	TRES INICIADOS	128	PENSAMENTO	PORTUGUÊS	BROCHURA	8531500710
2	O PODER DO HÁBITO	DUHIGG, CHARLES	408	OBJETIVA	PORTUGUÊS	BROCHURA	8539004119
3	QUEM PENSA ENRIQUECE	HILL, NAPOLEON	248	FUNDAMENTO	PORTUGUÊS	BROCHURA	8576766302
4	SOCIEDADE DO CANSAÇO	HAN, BYUNG-CHUL	136	VOZES	PORTUGUÊS	BROCHURA	8532649963
5	GESTALT-TERAPIA - FUNDAMENTOS EPISTEMOLOGICOS	FRAZAO, LILIAN MEYER	184	SUMMUS	PORTUGUÊS	BROCHURA	8532309089
6	INTRODUÇÃO A ALGEBRA LINEAR COM APLICAÇÕES	KOLMAN, BERNARD	684	LTC	PORTUGUÊS	BROCHURA	8521614780
7	O DIÁRIO DE ANNE FRANK	FRANK, ANNE	224	PRINCIPIS	PORTUGUÊS	BROCHURA	8594318286
8	O MUNDO SEGUNDO FELIPE NETO	NETO, FELIPE	128	PIXEL	PORTUGUÊS	CAPA	8555462096
9	O PEQUENO PRINCIPE (LIVRO DE BOLSO)	SAINT-EXUPERY, ANTOINE DE	96	CAMINHO	PORTUGUÊS	BROCHURA	8589987337
10	PRÓLOGO, ATO, EPÍLOGO	MONTENEGRO, FERNANDA	328	COMPANHIA	PORTUGUÊS	BROCHURA	8535932550
11	QUEM PENSA ENRIQUECE	HILL, NAPOLEON	248	FUNDAMENTO	PORTUGUÊS	BROCHURA	8576766302

Recall: 1.0
Precision: 1.0
F_Measure = 1.0
C = 12
N = 12
E = 12

WRAPPER DISAL

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	Ambiente E Os Processos De Maturacao, O	Winnicott, Dona	268	Artmed	Português	Livro	9788573074567
1	Autorretrato E Outras Crônicas	Andrade, Drummond De Andrade	256	Record	Português	Livro	9788501112330
2	Aventuras Na Netoland Com Luccas Neto	Neto, Luccas	64	Pixel Media (nova Fronteira)	Português	Livro	9788555461002
3	Exames Laboratoriais E Diagnosticos Em Enferma...	Fischbach, Frances	732	Guanabara (grupo Gen)	Português	Livro	9788527726634
4	Fisiologia Endocrina (Iange) - 4º Ed	Molina, Patricia E.	309	Mcgraw Hill (artmed)	Português	Livro	9788580553918
5	Gestao Hospitalar - Para Uma Administracao Eficaz	Malagon-londono, Gustavo	612	Guanabara (grupo Gen)	Português	Livro	9788527733298
6	Gramatica Da Lingua Portuguesa Para Leigos	Schlee, Magda Bahia	248	Alta Books	Português	Livro	9788550803395
7	Introducao A Analise Matematica	Avila, Geraldo	260	Edgard Blucher	Português	Livro	9788521201687
8	Michaelis Minidicionario Alemão - Alemão-portu...	Melhoramentos	480	Melhoramentos	Alemão-português / português- alemão	Livro	9788506078563
9	Michaelis Minidicionario Espanhol - Espanhol-p...	Melhoramentos	462	Melhoramentos	Português - Espanhol	Livro	9788506078532
10	Michaelis Minidicionario Frances - Frances-por...	Melhoramentos	510	Melhoramentos	Português - Francês	Livro	9788506078549

Recall: 1.0
 Precision: 1.0
 F_Measure = 1.0
 C = 11
 N = 11
 E = 11

WRAPPER AMERICANAS

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	A cinco passos de você	Rachael Lippincott	288	Globo Alt	Português	Livro brochura	9788525067425
1	A corrente	Adrian McKinty	378	Record	Português	Livro brochura	9788501117656
2	As 13 maldições (Vol. 2)	Michelle Harrison	434	Bertrand Brasil	Português	Livro brochura	9788528616668
3	Crônicas Saxônicas: A guerra do lobo (Vol. 11)	Bernard Cornwell	378	Record	Português	Livro brochura	9788501117359
4	Eu sou Eric Zimmerman - Volume 2	Megan Maxwell	400	Essência	Português	Livro brochura	9788542217209
5	Herdeiro de Sevenwaters - Vol. 4	Juliet Marillier	496	Butterfly	Português	Livro	9788568674055
6	História da filosofia cristã	Etienne Gilson, Philotheus Boehner	584	Editora Vozes	Português	Livro brochura	9788532607379
7	Medicina Dos Horrores: A história de Joseph Li...	Lindsey Fizharris	320	Intrinseca	Português	Livro Capa Dura	9788551005224
8	O cadete e o capitão	Luiz Maklouf Carvalho	256	Atual	Português	Livro brochura	9786580309351
9	Paixão libertadora	Sophie Jackson	352	Editora Arqueiro	Português	Livro brochura	9788580415421

Recall: 1.0
Precision: 1.0
F_Measure = 1.0
C = 11
N = 11
E = 11

WRAPPER SUBMARINO

	Título	Autor	Nº Páginas	Editora	Idioma	Acabamento	ISBN
0	Ambiente E Os Processos De Maturacao, O	Winnicott, Dona	268	Artmed	Português	Livro	9788573074567
1	Autorretrato E Outras Crônicas	Andrade, Drummond De Andrade	256	Record	Português	Livro	9788501112330
2	Aventuras Na Netoland Com Luccas Neto	Neto, Luccas	64	Pixel Media (nova Frenteira)	Português	Livro	9788555461002
3	Exames Laboratoriais E Diagnosticos Em Enferma...	Fischbach, Frances	732	Guanabara (grupo Gen)	Português	Livro	9788527726634
4	Fisiologia Endocrina (Iange) - 4º Ed	Molina, Patricia E.	309	Mcgraw Hill (artmed)	Português	Livro	9788580553918
5	Gestao Hospitalar - Para Uma Administracao Eficaz	Malagon-londono, Gustavo	612	Guanabara (grupo Gen)	Português	Livro	9788527733298
6	Gramatica Da Lingua Portuguesa Para Leigos	Schlee, Magda Bahia	248	Alta Books	Português	Livro	9788550803395
7	Introducao A Analise Matematica	Avila, Geraldo	260	Edgard Blucher	Português	Livro	9788521201687
8	Michaelis Minidicionario Alemao - Alemao-portu...	Melhoramentos	480	Melhoramentos	Alemão-português / português- alemão	Livro	9788506078563
9	Michaelis Minidicionario Espanhol - Espanhol-p...	Melhoramentos	462	Melhoramentos	Português - Espanhol	Livro	9788506078532
10	Michaelis Minidicionario Frances - Frances-por...	Melhoramentos	510	Melhoramentos	Português - Francês	Livro	9788506078549

Recall: 1.0
 Precision: 1.0
 F_Measure = 1.0
 C = 11
 N = 11
 E = 11

WRAPPER ÚNICO

88	História da filosofia cristã	Etienne Gilson, Philotheus Boehner	584	Editora Vozes	Português	Livro brochura	9788532607379
89	Medicina Dos Horrores: A história de Joseph Li...	Lindsey Fizharris	320	Intrinseca	Português	Livro Capa Dura	9788551005224
90	O cadete e o capitão	Luiz Maklouf Carvalho	256	Todavia	Português	Livro brochura	9786580309351
91	Paixão libertadora	Sophie Jackson	352	Editora Arqueiro	Português	Livro brochura	9788580415421
92	Luccas Neto em os aventureiros	Luccas Neto	64	Pixel	Português	Livro brochura	9788555461927
93	A dama mais apaixonada	Julia Quinn, Eloisa James, Connie Brockway	288	Editora Arqueiro	Português	Livro brochura	9788530600211
94	O Labirinto Do Fauno	Guillermo del Toro, Cornelia Funke	320	Intrinseca	Português	Livro Capa Dura	9788551005194
95	ABSOLUTE SANDMAN VOL. 1	Neil Gaiman	88	Panini Books	Português	Livro Capa Dura	9788573516531
96		Elizabeth Rudnick	224		Português	Livro	
97		Otomo, Katsuhiko	400		Português	Livro	
98		Veronica Roth	448		Português	Livro	9788579804557
99		Lauren Blakely	320		Português	Livro	9788595810846
100		Bráulio Bessa	160		Português	Livro	9788543108377
101	Seja Foda! - Black Edition	Caio Carneiro	208	Buzz Editora	Português	Livro Capa Dura	9786580435173

102 rows × 7 columns

Recall: 0.14705882352941177
Precision: 0.14705882352941177
F_Measure = 0.147
C = 15
N = 102
E = 102



WRAPPER PROBLEMAS

- As estrutura das páginas mudam muito
- Páginas dentro do mesmo domínio apresentam estruturas diferentes
 - Retorno de informação que não pertence àquela coluna
- Falta de dados dos produtos