

Querido diário

O projeto do Serenata de Amor para sua cidade!



OPEN KNOWLEDGE
BRASIL



José Guilherme Vanz

Embaixador de Inovação Cívica

Criando e resolvendo bugs desde
2010

Atualmente trabalhando com
Kubernetes e containers na SUSE
Linux

Organizador do Hackerspace
Blumenau

OPERAÇÃO

SERENATA DE AMOR

O QUE É SERENATA?

Inteligência artificial para engajamento cívico e controle social do governo

Ciência de dados usando as bases de dados abertos

Time multidisciplinar: cientistas sociais, programadores, estatísticos, publicitários, jornalistas

Open source: mais de 700 voluntários no grupo aberto do Telegram



COTA PARA EXERCÍCIO DE ATIVIDADE PARLAMENTAR - COMO FUNCIONA:



O parlamentar
realiza a compra
com o dinheiro
dele

Ele envia as **notas**
fiscais para a
Câmara dos
Deputados

O parlamentar é
reembolsado

TANTOS DADOS QUE PODEMOS ENSNINAR UM ROBÔ

CONHEÇAM A ROSIE!

INTELIGÊNCIA ARTIFICIAL → CONTROLE SOCIAL E ACCOUNTABILITY



Rosie

@RosieDaSerenata

A Robô da Operação Serenata de Amor. Analisa e identifica suspeitas em gastos de deputados federais em exercício de sua função.

© República Federativa do Brasil

Tweets **795** Following **572** Followers **37.2K** Likes **1**

Following

Tweets **Tweets & replies**



Rosie @RosieDaSerenata · 25 Dec 2018

 Gasto suspeito de Dep. @josededile (RS). Você pode me ajudar a verificar? jarbas.serenata.ai/layers/#/docum... #SerenataDeAmor na @CamaraDeputados

 Translate Tweet

 4  22  64 

Who to follow · Refresh · View all



Flávia Marreiro  @flavia... 

Follow



Maria Beltrão  @beltrao... 

Follow

**COMO ACESSAR E
ANALISAR AS
INFORMAÇÕES
PUBLICADAS NOS
DIÁRIOS OFICIAIS
DOS 5.570 MUNICÍPIOS**



QUERIDO DIÁRIO

Pesquisa [Pesquisa Avançada](#) [Menu](#)

Foram encontradas 43452 publicações para a seguinte pesquisa:



• Contendo os termos: **Prefeitura municipal de Blumenau**

Primeira Anterior 1 2 3 4 5 6 7 8 9 10 Próxima Última

CONCORRÊNCIA Nº 03-007/2020

27/04/2020 - Licitações - Prefeitura municipal de Blumenau - [\[Abrir/Salvar Original\]](#) - [\[Visualizar On-Line\]](#) - [\[Abrir Edição\]](#) - [\[Imprimir Extrato\]](#)

-mail: licitacoes@blumenau.sc.gov.br e/ou site oficial do Município <http://www.blumenau.sc.gov.br/transparencia>. BASE LEGAL: Lei nº 8.666/93 e alterações e legislação complementar. Informações (47) 3381-6277/(47) 3381-6724. **Blumenau**, 27/04/2020. ANDERSON ROSA - SECRETÁRIO MUNICIPAL DE ADMINISTRAÇÃO.

[Resultados Semelhantes...](#)

CONCORRÊNCIA Nº 03-008/2020

27/04/2020 - Licitações - Prefeitura municipal de Blumenau - [\[Abrir/Salvar Original\]](#) - [\[Visualizar On-Line\]](#) - [\[Abrir Edição\]](#) - [\[Imprimir Extrato\]](#)

-mail: licitacoes@blumenau.sc.gov.br e/ou site oficial do Município <http://www.blumenau.sc.gov.br/transparencia>. BASE LEGAL: Lei nº 8.666/93 e alterações e legislação complementar. Informações (47) 3381-6277/(47) 3381-6724. **Blumenau**, 27/04/2020. ANDERSON ROSA - SECRETÁRIO MUNICIPAL DE ADMINISTRAÇÃO.

[Resultados Semelhantes...](#)

TERMO DE HOMOLOGAÇÃO CONCORRÊNCIA Nº 03-2208/2019 - SAMAE

27/04/2020 - Licitações - Prefeitura municipal de Blumenau - [\[Abrir/Salvar Original\]](#) - [\[Visualizar On-Line\]](#) - [\[Abrir Edição\]](#) - [\[Imprimir Extrato\]](#)

seu resultado final com base nos documentos acostados no processo, sendo: 1º lugar: Samil Terraplanagens e Transportes EIRELI - EPP. Valor global de R\$ 6.951.591,63 (seis milhões, novecentos e cinquenta e um mil e quinhentos e noventa e um reais e sessenta e três centavos). ADJUDICAÇÃO Pelas razões acima expostas ADJUDICO o objeto da licitação modalidade Conc. 03-2208-2019. BASE LEGAL: Lei 8.666/93 de 21 de junho de 1993 e suas alterações. **Blumenau**, 24 de abril de 2020 Michael R. Schneider

[Resultados Semelhantes...](#)

COMO É UM DIÁRIO OFICIAL ?



QUERIDO DIÁRIO

CONCORRENCIA 03-2208/2019

Objeto: A presente licitação tem como objeto a locação de máquinas e equipamentos para prestação de serviços de terraplanagem em geral, transporte e carregamento de materiais diversos para reaterro e bota-fora; conforme necessidades do SAMAE, de acordo com especificações, quantitativos e condições estabelecidas no presente edital e anexos.

HOMOLOGAÇÃO

Apreciada a classificação final da licitação modalidade Concorrência, com o objeto acima mencionado, **HOMOLOGO** seu resultado final com base nos documentos acostados no processo, sendo:

1º lugar: **Samil Terraplanagens e Transportes EIRELI - EPP**. Valor global de **RS 6.951.591,63** (seis milhões, novecentos e cinquenta e um mil e quinhentos e noventa e um reais e sessenta e três centavos).

ADJUDICAÇÃO

Pelas razões acima expostas **ADJUDICO** o objeto da licitação modalidade Conc. 03-2208-2019.

BASE LEGAL: Lei 8.666/93 de 21 de junho de 1993 e suas alterações.

Blumenau, 24 de abril de 2020

COMO É UM **DIÁRIO** **OFICIAL ?**



QUERIDO **DIÁRIO**

ACESSA O SITE DA PREFEITURA

ENCONTRA O DIÁRIOS OFICIAIS

BAIXA OS ARQUIVOS

TRANSFORMA ARQUIVOS EM
TEXTO PURO

SALVA ARQUIVO TEXTO E
ORIGINAL NA MÁQUINA. SALVA
NO BANCO DE DADOS

COMO OBTEMOS OS
DIÁRIOS OFICIAIS ?

**AUTOMATIZAMOS
PESSOAS! ;-)**



QUERIDO DIÁRIO

```
jvanz@earth:~/serenata/diario-oficial> ls processing/data_collection/gazette/spiders/  
al_maceio.py          sc_faxinal_dos_guedes.py    sc_pescaria_brava.py  
am_manaus.py         sc_flor_do_sertao.py       sc_petrolandia.py  
ba_feira_de_santana.py sc_florianopolis.py       sc_pinhalzinho.py  
ba_salvador.py       sc_formosa_do_sul.py      sc_pinheiro_preto.py  
base.py              sc_forquilha.py           sc_piratuba.py  
ce_fortaleza.py     sc_fraiburgo.py           sc_planalto_alegre.py  
es_associacao_municipios.py sc_frei_rogerio.py       sc_pomerode.py  
go_aparecida_de_goiania.py sc_galvao.py              sc_ponte_alta_do_norte.py  
go_goiania.py        sc_garopaba.py            sc_ponte_serrada.py  
__init__.py          sc_garuva.py               sc_porto_belo.py  
mg_uberaba.py        sc_gaspar.py               sc_porto_uniao.py  
ms_campo_grande.py  sc_governador_celso_ramos.py sc_pouso_redondo.py  
pr_cascavel.py      sc_grao_para.py           sc_praia_grande.py  
pr_curitiba.py      sc_gravatal.py            sc_presidente_castello_branco.py  
pr_foz_do_iguacu.py sc_guabiruba.py           sc_presidente_getulio.py  
pr_londrina.py      sc_guaraciaba.py          sc_presidente_nereu.py  
pr_maringa.py       sc_guaramirim.py          sc_princesa.py  
pr_ponta_grossa.py  sc_guaruja_do_sul.py      sc_quilombo.py
```

303 SPIDERS

**MAS COMO ISSO
FUNCIONA NA PRÁTICA ?**

**PRECISAMOS CRIAR UM
ROBOZINHO PARA CADA
CIDADE.**



QUERIDO DIÁRIO


```

50
51 class ExtractTextPipeline:
52     """
53     Identify file format and call the right tool to extract the text from it
54     """
55
56     def process_item(self, item, spider):
57         if self.is_doc(item["files"][0]["path"]):
58             item["source_text"] = self.doc_source_text(item)
59         elif self.is_pdf(item["files"][0]["path"]):
60             item["source_text"] = self.pdf_source_text(item)
61         elif self.is_txt(item["files"][0]["path"]):
62             item["source_text"] = self.txt_source_text(item)
63         else:
64             raise Exception(
65                 "Unsupported file type: " + self.get_file_type(item["files"][0]["path"])
66             )
67
68         for key, value in item["files"][0].items():
69             item[f"file_{key}"] = value
70         item.pop("files")
71         item.pop("file_urls")
72         return item
73
74     def pdf_source_text(self, item):
75         """
76         Gets the text from pdf files
77         """
78         pdf_path = os.path.join(FILES_STORE, item["files"][0]["path"])
79         text_path = pdf_path + ".txt"
80         command = f"pdftotext -layout {pdf_path} {text_path}"
81         subprocess.run(command, shell=True, check=True)
82         with open(text_path) as file:
83             return file.read()
84
85     def doc_source_text(self, item):
86         """
87         Gets the text from docish files
88         """
89         doc_path = os.path.join(FILES_STORE, item["files"][0]["path"])
90         text_path = doc_path + ".txt"
91         command = f"java -jar /tika-app.jar --text {doc_path}"
92         with open(text_path, "w") as f:
93             subprocess.run(command, shell=True, check=True, stdout=f)
94         with open(text_path, "r") as f:
95             return f.read()
96
97     def txt_source_text(self, item):
98         """
99         Gets the text from txt files
100         """
101         with open(
102             os.path.join(FILES_STORE, item["files"][0]["path"]), encoding="ISO-8859-1"
103         ) as f:
104             return f.read()
105
106     def is_pdf(self, filepath):
107         """
108         If the file type is pdf returns True. Otherwise,
109         returns False
110         """
111         return self._is_file_type(filepath, file_types=["application/pdf"])
112
113     def is_doc(self, filepath):
114         """
115         If the file type is doc or similar returns True. Otherwise,
116         returns False
117         """
118         file_types = [
119             "application/msword",
120             "application/vnd.oasis.opendocument.text",
121             "application/vnd.openxmlformats-officedocument.wordprocessingml.document",
122         ]
123         return self._is_file_type(filepath, file_types)
124
125     def is_txt(self, filepath):
126         """
127         If the file type is txt returns True. Otherwise,
128         returns False
129         """
130         return self._is_file_type(filepath, file_types=["text/plain"])
131

```

https://github.com/okfn-brasil/diario-oficial/blob/master/processing/data_collection/gazette/pipelines.py

jvanz@earth:~/serenata/diario-oficial> ls data/full/
000468c3df9db8e0aacd685679d93c48b4800c7f.doc.txt
000468c3df9db8e0aacd685679d93c48b4800c7f.doc.txt
0008bc8f5fd6e4944ad10f8200e54ba7cc3750ef.doc.txt
0008bc8f5fd6e4944ad10f8200e54ba7cc3750ef.doc.txt
001145a82d39f0d554de6672361bf626603304503.doc.txt
001145a82d39f0d554de6672361bf626603304503.doc.txt
002507e999037719f7f4016ca518e57df42b8fec.pdf.txt
002507e999037719f7f4016ca518e57df42b8fec.pdf.txt
002653cd3f1d9e9a075998cde08c28198f0cb803.doc.txt
002653cd3f1d9e9a075998cde08c28198f0cb803.doc.txt
003932d068dc2702f77d22c8f207c4f30c699696.doc
003932d068dc2702f77d22c8f207c4f30c699696.doc.txt
00412d6e1e3717738792a0997387ebc284f66be6.doc
00412d6e1e3717738792a0997387ebc284f66be6.doc.txt
005104b734a9b03d653d89e86c39e44166dd915.doc.txt
005104b734a9b03d653d89e86c39e44166dd915.doc.txt
0057f89b11488bf9c8d65746e16365b1e6bbe6de6.doc
0057f89b11488bf9c8d65746e16365b1e6bbe6de6.doc.txt
005ed65cc85f0f1fa951115f3591ceb9cb1547ed.pdf
005ed65cc85f0f1fa951115f3591ceb9cb1547ed.pdf.txt
006ba97ad4ecb1086cd349f70133a5427356a183.pdf
006ba97ad4ecb1086cd349f70133a5427356a183.pdf.txt
006c1d46edf8cd1a52a1835c73005b0d00e8062.doc
006c1d46edf8cd1a52a1835c73005b0d00e8062.doc.txt
006c879ffabb1d5b862145c86c881d6e081312c6.doc
006c879ffabb1d5b862145c86c881d6e081312c6.doc.txt
007a4bf262d3068c6a34138d4ebd55b65c28f198.doc
007a4bf262d3068c6a34138d4ebd55b65c28f198.doc.txt
007b5818d9e44ed54801f491beab63fa8c77c71e.pdf
007b5818d9e44ed54801f491beab63fa8c77c71e.pdf.txt
00985c3abf0c16706e6662e6e130b3c93c7559ec.doc
00985c3abf0c16706e6662e6e130b3c93c7559ec.doc.txt
0099ae004ce02b597eddbae3ae312bad6baed29a.doc
0099ae004ce02b597eddbae3ae312bad6baed29a.doc.txt
009ee33f2b20f16b70035d5eabe103ed52a3ec1f.doc
009ee33f2b20f16b70035d5eabe103ed52a3ec1f.doc.txt
00ad50efffa31a0f0deb4869e54cda2eaa66cc2.doc
00ad50efffa31a0f0deb4869e54cda2eaa66cc2.doc.txt
00b4db6a3363a38cda12f25277ce73d09f2d189.pdf
00b4db6a3363a38cda12f25277ce73d09f2d189.pdf.txt
00bc96f156dc0e648d13025b92c1a77d098f27432.doc
00bc96f156dc0e648d13025b92c1a77d098f27432.doc.txt
00d8496ef5b06543d5436ea667579e2106934eb5.doc
00d8496ef5b06543d5436ea667579e2106934eb5.doc.txt
00dd086d34a3632edeb94cfca6f6dabed6caca.docx
00dd086d34a3632edeb94cfca6f6dabed6caca.docx.txt
00dde2980e74b84ac27d81d9c277f943e827d77b.pdf
00dde2980e74b84ac27d81d9c277f943e827d77b.pdf.txt
00f1c4506dad1b16eff855a861f1f22a1e2644e2.doc
00f1c4506dad1b16eff855a861f1f22a1e2644e2.doc.txt
00f4402ec5c056eb1b64f02362bd92d344dea506.pdf
00f4402ec5c056eb1b64f02362bd92d344dea506.pdf.txt
0102a899319ebd20c3a3fa28a8f6b276a2d02b6d.doc
0102a899319ebd20c3a3fa28a8f6b276a2d02b6d.doc.txt
0122d46b4c38ba058e622e9fa6c67f8029b0480.doc
0122d46b4c38ba058e622e9fa6c67f8029b0480.doc.txt
01243a35b1dd442917148deF574b307304e06e34.doc

O QUE FAZEMOS COM DIÁRIOS OFICIAIS ?

ATUALMENTE... NADA.

T_T



QUERIDO DIÁRIO

okfn-brasil / diario-oficial

Unwatch 64 Unstar 425 Fork 113

Code Issues 14 Pull requests 31 Actions Security 0 Insights Settings

Labels Milestones

Edit milestone


New issue

0.9.0

No due date 55% complete

Milestone to keep track of the tasks necessary to adjust the project to run in a production environment

4 Open 5 Closed

- Production infrastrucure configuration**  2
#160 opened on Mar 2 by jvanz
- Make the gazettes available for the public** Epic 2
#157 opened on Feb 20 by jvanz
- WebUI to find and visualize the gazettes** Epic 2
#161 opened on Mar 3 by jvanz
- Upload scraped files to digital ocean** 2
#165 opened on Mar 7 by jvanz

Tip! You can use `shift+j` or `shift+k` to move items with your keyboard.

O QUE GOSTARÍAMOS DE FAZER ?

PRECISAMOS DE VOCÊ!



QUERIDO DIÁRIO

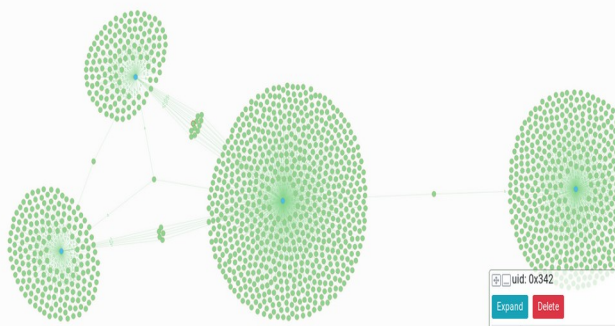
localhost

Console

Query Mutate Read Only Best Effort Clear Run

```
1 {
2   all(func: has(estado)){
3     uid
4     nome
5     type
6     estado
7     business {
8       cnpj
9       type
10      name
11    }
12  }
13 }
14
```

Graph JSON Request



uid: 0x342

Expand Delete

pred.	value
cnpj	"33645482000196"
name	"INSTITUTO BRASILEIRO DE ADMINISTRACAO MUNICIPAL IBAM"
type	"CNPJ"
uid	"0x342"

business b

HISTORY

q { all(func: has(estado))(uid nome type estado business (...)) }

q { all(func: eq(nome, "Blumenau"))(uid nome type estado business (...)) }

PODERIA SER FEITO
MAIS **NO FUTURO...**



QUERIDO DIÁRIO

Search or jump to... Pull requests Issues Marketplace Explore

okfn-brasil / diario-oficial Unwatch 64 Unstar 425 Fork 113

Code Issues 15 Pull requests 31 Actions Security 0 Insights Settings

Brazilian government gazettes, accessible to everyone. Edit

Manage topics

684 commits 5 branches 0 packages 0 releases 30 contributors MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

jvanz Merge pull request #166 from AndreAngelucci/master Latest commit 1a2b35e 19 days ago	
.github/workflows	build and publish container image 2 months ago
data	Revert "Fix permission denied errors on running from scratch" 2 years ago
processing	Update Apache Tika to 1.24 in Dockerfile 20 days ago
.env.example	Remove code related to bidding exemption parsing 11 months ago
.gitignore	scrapy-deltafetch: avoid crawl the same URL 2 months ago
.pre-commit-config.yaml	Upgrade dependencies 11 months ago
.travis.yml	set travis to python 3.7 5 months ago
CITIES.md	Update CITIES.md with RS Gravatal updates 2 months ago
CODE_OF_CONDUCT.md	Add code of conduct 2 years ago
CONTRIBUTING.md	Makefile: helper commands 5 months ago
LICENSE	Transfer copyright to Open Knowledge Foundation Brasil 2 years ago
Makefile	build and publish container image 2 months ago
README.md	Add more instructions to execute the spiders 5 months ago
docker-compose.yml	Remove code related to bidding exemption parsing 11 months ago

README.md

Diário Oficial

Diário Oficial is the Brazilian government gazette, one of the best places to know the latest actions of the public administration, with distinct publications in the federal, state and municipal levels.

Even with recurrent efforts of enforcing the [Freedom of Information legislation](#) across the country, official communication remains - in most of the territories - in PDFs.

The goal of this same project is to upgrade **Diário Oficial** to the year when it gets published, making accessible information that is currently only available.

When this project was initially released, had two distinct goals: creating crawlers for governments gazettes and parsing bidding exemptions from them. Now going forward, it is limited to the first objective.

NÃO SE ACANHE. NOS
PROCURE!

<https://github.com/okfn-brasil/diario-oficial>



Obrigado!