

ÂNDREI ARTAXEXES DE JESUS FERREIRA

*WEB SCRAPER* PARA O DIÁRIO OFICIAL  
DA UNIÃO

Taguatinga, DF

2016

ÂNDREI ARTAXEXES DE JESUS FERREIRA

# *WEB SCRAPER* PARA O DIÁRIO OFICIAL DA UNIÃO

Trabalho de Conclusão de Curso apresentado à banca examinadora do Centro Universitário Estácio de Brasília, como um dos requisitos para a obtenção do grau de Bacharel em Sistemas de Informação.

Centro Universitário Estácio de Brasília

Orientador: Prof. Esp. João Paulo Pimentel

Taguatinga, DF

2016

ÁNDREI ARTAXEXES DE JESUS FERREIRA

*WEB SCRAPER* PARA O DIÁRIO OFICIAL DA UNIÃO

Trabalho de Conclusão de Curso aprovado  
como requisito parcial para obtenção de grau  
de Bacharel em Sistemas de Informação pelo  
Centro Universitário Estácio de Brasília.

Banca examinadora:

Taguatinga, DF, 2016

---

Prof. Esp. João Paulo Pimentel  
Centro Universitário Estácio de Brasília  
Presidente (Orientador)

---

Prof. Esp. André Luis Gomes Pompas  
Centro Universitário Estácio de Brasília  
1º Examinador

---

Prof. Esp. Jósias Alves de Souza Filho  
Centro Universitário Estácio de Brasília  
2º Examinador

# Agradecimentos

Aos professores: pela profissão que escolheram e pelo conhecimento compartilhado.

Às amigadas, próximas ou não: pelo companheirismo e preocupação. Aqui são consideradas as de classe (Ensino Fundamental, Médio e Superior), pelas vivências e irmandades; as pessoais, pelo afeto e dedicação; e as profissionais, pelas experiências compartilhadas e pelos cafés!

Aos familiares: pelo suporte incondicional.

Ao carinho e compreensão da namorada.

E principalmente, aos pais: pela criação, sacrifício e apoio; sempre indicando o que acreditam ser o caminho correto a se seguir.

# Lista de ilustrações

Figura 1 – Diagrama de Caso de Uso . . . . .	18
--	----

# Lista de tabelas

Tabela 1 – Cronograma previsto . . . . .	11
Tabela 2 – Cronograma realizado . . . . .	11
Tabela 3 – Recursos utilizados - <i>Hardware</i> . . . . .	12
Tabela 4 – Recursos utilizados - <i>Software</i> . . . . .	12
Tabela 5 – Recursos utilizados - Humano . . . . .	12
Tabela 6 – <i>Top Analytics/Data Science Tools</i> . . . . .	13
Tabela 7 – <i>DB-Engines Ranking of Search Engines</i> . . . . .	15
Tabela 8 – Caso de Uso - Baixar DOU para disco local . . . . .	19
Tabela 9 – Caso de Uso - Baixar DOU para <i>NoSQL</i> . . . . .	19
Tabela 10 – Caso de Uso - Baixar DOU para <i>search engine</i> . . . . .	20

# Lista de abreviaturas e siglas

DOU	Diário Oficial da União
<i>GUI</i>	<i>Graphical User Interface</i>
<i>HTTP</i>	<i>Hypertext Transfer Protocol</i>
<i>JSON</i>	<i>JavaScript Object Notation</i>
<i>KNIME</i>	<i>Konstanz Information Miner</i>
<i>NoSQL</i>	<i>Not Only Structured Query Language</i>
<i>PaaS</i>	<i>Platform as a Service</i>
<i>PSF</i>	<i>Python Software Foundation</i>
<i>RESTful</i>	<i>Representational State Transfer</i>
<i>RDBMS</i>	<i>Relational Database Management System</i>
SGBD	Sistema de Gerenciamento de Banco de Dados
<i>SQL</i>	<i>Structured Query Language</i>
<i>UML</i>	<i>Unified Modeling Language</i>
<i>VI</i>	<i>Visual Editor</i>
<i>VIM</i>	<i>Visual Editor Improved</i>

# Sumário

<b>1</b>	<b>Introdução</b>	<b>9</b>
1.1	Apresentação	9
1.2	Formulação do problema	9
1.3	Justificativa	9
1.4	Objetivos	10
1.4.1	Objetivo Geral	10
1.4.2	Objetivos Específicos	10
1.5	Definição do Escopo	10
1.6	O Mercado	10
1.7	Metodologia	11
1.8	Cronograma	11
1.8.1	Previsto	11
1.8.2	Realizado	11
1.9	Recursos Utilizados	12
1.9.1	<i>Hardware</i>	12
1.9.2	<i>Software</i>	12
1.9.3	Humano	12
<b>2</b>	<b>Ferramentas e tecnologias utilizadas</b>	<b>13</b>
2.1	<i>Python</i>	13
2.2	<i>VIM</i>	14
2.3	<i>MongoDB</i>	14
2.4	<i>Elasticsearch</i>	15
2.5	<i>Web Scraping</i>	15
2.6	Diário Oficial da União	16
2.7	<i>Astah Community</i>	16
2.8	<i>UML</i>	16
<b>3</b>	<b>O desenvolvimento</b>	<b>17</b>
3.1	Análise de Requisitos	17
3.1.1	Funcionais	17
3.1.2	Não funcionais	17
3.2	Caso de Uso	17
3.2.1	Visão geral do Caso de Uso	17
3.2.2	Diagrama de Caso de Uso	18
3.2.3	Especificação dos Casos de Uso	19
3.2.3.1	Caso de Uso Baixar DOU para disco local	19
3.2.3.2	Caso de Uso Baixar DOU para <i>NoSQL</i>	19



3.2.3.3	Caso de Uso Baixar DOU para <i>search engine</i>	20
3.3	Diagrama de Projeto	20
3.3.1	Especificação das Classes	20
3.4	Diagramas de Sequência	20
3.5	Diagrama de Estado	20
3.6	Interface Visual do Sistema	20
<b>4</b>	<b>Conclusão e trabalhos futuros</b>	<b>21</b>
4.1	Conclusão	21
4.2	Trabalhos futuros	21
	<b>Referências</b>	<b>22</b>

# 1 Introdução

Em um cenário onde são explorados cada vez mais os conceitos de *Big Data*, se faz necessário o uso de ferramentas alternativas para a extração de dados semi-estruturados e não estruturados das mais variadas fontes, observando-se assim uma das características dessa tecnologia: a variedade dos tipos de dados.

Considerando a necessidade de *web scrapers* específicos para nichos de dados, é proposto nesse trabalho o desenvolvimento de um *software* especificamente para acesso ao conteúdo divulgado no Diário Oficial da União (DOU), um dos veículos de comunicação do Governo Federal pelo qual são publicados todos os seus atos.

## 1.1 Apresentação

Esse trabalho objetiva o desenvolvimento de um *software*, nomeado *scrpr4dou* (lê-se: *scraper for DOU*), para execução em Sistemas Operacionais suportados pela linguagem de programação *Python*, o qual auxiliará o *Data Engineer* e/ou *Data Scientist* no processo de alimentação de um *Data Lake* com arquivos provenientes do Diário Oficial da União, possibilitando seu armazenamento em bancos de dados *Not Only SQL (NoSQL)* e sua exploração por mecanismos de busca.

## 1.2 Formulação do problema

A variedade dos tipos de dados existentes na *Internet* não é novidade, mas a necessidade de se automatizar o processo de obtenção desses dados cresceu com o advento do termo *Big Data*.

Diante de tal necessidade, a técnica de *web scraping*, que busca extrair informações de *websites*, se tornou cada vez mais comum. Existem *softwares* que atendem essa necessidade de forma satisfatória, mas em um contexto generalizado.

O Diário Oficial da União passa a ser um caso específico quando não interliga seus arquivos publicados de forma a atender ou facilitar o *modus operandi* desses *softwares* generalistas.

## 1.3 Justificativa

Tornar os dados e informações contidos no Diário Oficial da União mais acessíveis aos dispositivos com capacidade computacional, para que sejam passíveis de análise

automatizada, tornando assim a visualização e exploração do DOU pelo usuário mais simples e enriquecedora.

## 1.4 Objetivos

### 1.4.1 Objetivo Geral

Disponibilizar à comunidade um *software* baseado em *Python* para permitir e facilitar o processo de *web scraping* nos arquivos publicados pelo Diário Oficial da União.

### 1.4.2 Objetivos Específicos

Realizar o levantamento de regras permissivas ou não, relacionadas ao processo de *web crawler* do conteúdo disponibilizado pela Imprensa Nacional.

Realizar estudo sobre o atual modo de indexação dos arquivos alvos.

Utilizar bibliotecas *Hypertext Transfer Protocol* (HTTP) disponíveis pela linguagem de programação *Python* para executar o *download* dos arquivos de maneira automatizada.

Realizar a integração da ferramenta desenvolvida com *softwares* utilizados pela comunidade de *Big Data*.

## 1.5 Definição do Escopo

O foco do presente projeto é a facilitação do acesso ao conteúdo do DOU por ferramentas que processem dados semi ou não estruturados.

Este projeto não engloba, num primeiro momento, a visualização gráfica do conteúdo nem a percepção de *insights* a partir de seus dados, tornando essas funcionalidades parte do escopo de trabalhos futuros por se tratar de continuidade do atual projeto.

Este *web scraper* se limita a armazenar e permitir a interação com o conteúdo baixado, não incorporando funções clássicas de *web crawlers*, *web robots* ou *spiders*, tais como *injector*, *updater* ou *link inverter*, dentre outros.

## 1.6 O Mercado

A ferramenta desenvolvida atende a necessidade de órgãos públicos que precisam de alguma forma extrair dados do Diário Oficial da União, além de empresas interessadas em desenvolver produtos para o mesmo.

## 1.7 Metodologia

Pesquisa bibliográfica, em *websites* e em documentação das ferramentas utilizadas, desenvolvimento em linguagem de programação *Python* e seus pacotes, integração com banco de dados *NoSQL MongoDB* e *Elasticsearch*, além de recursos de apoio tais como a *Unified Modeling Language (UML)* e o *Astah Community*.

## 1.8 Cronograma

### 1.8.1 Previsto

Tabela 1: Cronograma previsto

Etapas	AGO/2016				SET/2016				OUT/2016				NOV/2016				
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5 <sup>a</sup>
Tema																	
Leitura/estudo																	
Pré-projeto																	
Requisitos																	
Documentação																	
Codificação																	
Teste																	
Revisão																	
Apresentação																	

Fonte: Pesquisa de campo (2016)

### 1.8.2 Realizado

Tabela 2: Cronograma realizado

Etapas	AGO/2016				SET/2016				OUT/2016				NOV/2016				
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5 <sup>a</sup>
Tema																	
Leitura/estudo																	
Pré-projeto																	
Requisitos																	
Documentação																	
Codificação																	
Teste																	
Revisão																	
Apresentação																	

Fonte: Pesquisa de campo (2016)

## 1.9 Recursos Utilizados

### 1.9.1 Hardware

Tabela 3: Recursos utilizados - *Hardware*

<b><i>Hardwares</i></b>	<b>Valor</b>
1 <i>Netbook</i>	R\$ 1.200,00
1 <i>Macbook</i>	R\$ 100,00
1 <i>Desktop</i>	R\$ 2.400,00

Fonte: Pesquisa de campo (2016)

### 1.9.2 Software

Tabela 4: Recursos utilizados - *Software*

<b><i>Softwares</i></b>	<b>Valor</b>
<i>Python</i>	Gratuito
Banco de dados <i>MongoDB</i>	Gratuito
<i>Search engine Elasticsearch</i>	Gratuito
<i>VIM</i>	Gratuito
<i>Astah Community</i>	Gratuito
<i>UML</i>	Gratuito

Fonte: Pesquisa de campo (2016)

### 1.9.3 Humano

Tabela 5: Recursos utilizados - Humano

<b>Nome</b>
Anddrei Artaxexes de Jesus Ferreira

Fonte: Pesquisa de campo (2016)

## 2 Ferramentas e tecnologias utilizadas

### 2.1 *Python*

*Python* é uma linguagem de programação que teve seu desenvolvimento iniciado em 1990 por Guido van Rossum, sendo caracterizada por sua estabilidade e maturidade, além de possuir propósito geral e ser de muito auto-nível, dinâmica, orientada à objetos e multiplataforma (MARTELLI, 2006, p. 3).

A *Python Software Foundation (PSF)*, organização sem fins lucrativos detentora dos direitos autorais do *Python*, que tem por missão promover, proteger e desenvolver a linguagem, afirma que *Python* pode ser aplicado em soluções para vários tipos diferentes de problemas, o que expande sua adoção em vários ambientes dentre os suportados para solucionar demandas diversas (PSF, 2016).

*Python* utiliza tipagem dinâmica e proporciona boa leitura de seu código-fonte por sua sintaxe clara, e atualmente possui duas versões em desenvolvimento e/ou manutenção concorrente: a 2.x e a 3.x. As versões 3.x não são compatíveis com as versões 2.x, pois a ideia ao se criar a versão 3.x foi justamente se distanciar de algumas características peculiares da versão 2.x (DRISCOLL, 2014, p. 3).

*Python* foi escolhida para o desenvolvimento deste trabalho por estar entre as ferramentas mais populares de *Data Science*, segundo indica pesquisa realizada em junho de 2016, em comparação com o ano anterior, sobre as ferramentas de *Analytics/Data Science* mais utilizadas, conforme demonstrado na Tabela 6, além de ser uma linguagem de *script* e multiplataforma.

Tabela 6: *Top Analytics/Data Science Tools*

<i>Tool</i>	<b>2016 % share</b>	<b>% share</b>	<b>% alone</b>
R	49.0%	+4.5%	1.4%
<i>Python</i>	45.8%	+51.0%	0.1%
<i>SQL</i>	35.5%	+15.0%	0%
<i>Excel</i>	33.6%	+47.0%	0.2%
<i>RapidMiner</i>	32.6%	+3.5%	11.7%
<i>Hadoop</i>	22.1%	+20.0%	0%
<i>Spark</i>	21.6%	+91.0%	0.2%
<i>Tableau</i>	18.5%	+49.0%	0.2%
<i>KNIME</i>	18.0%	-10.0%	4.4%
<i>scikit-learn</i>	17.2%	+107%	0%

Fonte: *KDnuggets 2016 Software Poll Results* (2016)

## 2.2 VIM

VIM, acrônimo para *VI Improved*, é o clone mais popular e mais utilizado do *Visual Editor* (VI), editor de texto disponível em todos os sistemas *Unix* modernos), sendo escrito e mantido por Bram Moolenaar, e continua crescendo em termos de funcionalidade (ROBBINS; HANNAH; LAMB, 2008).

## 2.3 MongoDB

Segundo Banker et al. (2016, p. 5), "o 'termo guarda-chuva' *NoSQL* foi criado por volta de 2009 para agrupar os muitos bancos de dados não-relacionais populares da época, tendo em comum o uso de outra linguagem de consulta diferente de *SQL*".

Banker et al. (2016, p. 4) ainda afirma que no *MongoDB* "o modelo de dados e as estratégias de persistências são construídos para operações de leitura e escrita com alta taxa de transferência e fácil capacidade de escalonamento com tolerância à falhas".

Essas características são muito importantes em aplicações de *Big Data*, pois a demanda por volume de dados e sua velocidade e variedade, além de outras características, podem ser atendidas.

*MongoDB* é um banco de dados não-relacional orientado a documentos *JSON like* desenvolvido em 2007 como parte de uma Plataforma como Serviço (*PaaS*).

Enquanto bancos de dados relacionais (*RDBMS*) necessitam de tabelas esquematizadas, o *MongoDB* armazena as informações em documentos no lugar de colunas, livres de esquemas e sem necessidade de conhecimento prévio do dado a ser armazenado. Enquanto *RDBMSs* possuem tabelas, *MongoDB* possui *collections*, que representa um grupo de documentos.

Exemplo de documento no *MongoDB*:

```
{
  "_id": 1,
  "author": { "name" : "anddrei ferreira", "username": "artaxexes" },
  "text": "let's be a data scientist",
  "date": "2017-01-23 14:00UTC",
  "location": [ 37.7918781, -122.3692244 ],
  "comments": [
    { "username": "girlfriend", "text": "let's graduate" }
  ]
}
```

## 2.4 *Elasticsearch*

É necessário lidar com alguns problemas para tornar um conjunto de dados pesquisável, dentre os quais pode-se destacar o retorno de resultados relevantes e o retorno de estatísticas. *Search engines*, ou motores de busca, tais como *Elasticsearch*, atendem esses desafios (GHEORGHE; HINMAN; RUSSO, 2015).

*Elasticsearch* é uma *engine open-source RESTful* de pesquisa e análise que utiliza *Apache Lucene*, uma biblioteca de mecanismos de busca também *open-source*.

*Elasticsearch* oferece a combinação de uma poderosa *full-text search engine* com *data analytics* em larga escala, e a interação se dá com o uso do formato *JSON* utilizando-se requisições com verbos do protocolo *HTTP* ao servidor.

Considerando-se somente *search engines*, o *Elasticsearch* atualmente figura em primeiro lugar no ranking *DB-Engines*, como mostra a tabela abaixo:

Tabela 7: *DB-Engines Ranking of Search Engines*

<b>Rank</b>	<b>DBMS</b>	<b>Database Model</b>
1	<i>Elasticsearch</i>	<i>Search engine</i>
2	<i>Solr</i>	<i>Search engine</i>
3	<i>Splunk</i>	<i>Search engine</i>
4	<i>MarkLogic</i>	<i>Multi-model</i>
5	<i>Sphinx</i>	<i>Search engine</i>
6	<i>Google Search Appliance</i>	<i>Search engine</i>
7	<i>Amazon CloudSearch</i>	<i>Search engine</i>
8	<i>Microsoft Azure Search</i>	<i>Search engine</i>
9	<i>Algolia</i>	<i>Search engine</i>
10	<i>Xapian</i>	<i>Search engine</i>

Fonte: *DB-Engines Ranking of Search Engines* (NOV/2016)

*DB-Engines Ranking*, *ranking* criado e mantido pela empresa austríaca de consultoria em Tecnologia da Informação *Solid IT*, é uma lista de sistemas de gerenciamento de banco de dados (SGBD) baseada em popularidade e atualizada mensalmente. Dentre os critérios adotados pelo *ranking* estão número de menções em *websites*, número de empregos ofertados com menção ao SGBD, relevância nas redes sociais, etc (DB-ENGINE, 2016).

## 2.5 *Web Scraping*

É uma técnica utilizada para extração de dados de *websites* na qual o *software* acessa a *World Wide Web* utilizando protocolos de transferência, como por exemplo o *Hypertext Transfer Protocol*.



## 2.6 Diário Oficial da União

O Diário Oficial da União é o jornal oficial do governo federal desde 1862 pelo qual torna-se público seus atos.

É publicado pela Imprensa Nacional, um dos órgãos mais antigos do país.

## 2.7 *Astah Community*

Versão gratuita da ferramenta de *design* de *software* *Astah Professional*, que permite a criação de diagramas *UML*.

## 2.8 *UML*

Acrônimo para *Unified Modeling Language*, linguagem desenvolvida para padronizar a visualização do *design* de um *software*.

## 3 O desenvolvimento

### 3.1 Análise de Requisitos

#### 3.1.1 Funcionais

O usuário deverá poder executar o *download* automatizado dos arquivos publicados pela Imprensa Nacional referentes ao Diário Oficial da União.

O *software* deverá possibilitar o armazenamento dos arquivos baixados no sistema de arquivos local.

O *software* deverá permitir o armazenamento do conteúdo baixado do Diário Oficial da União em banco de dados não-relacional.

O *software* deverá possibilitar o uso do conteúdo dos arquivos baixados em uma *engine* de pesquisa textual.

#### 3.1.2 Não funcionais

O *software* deverá ser executável principalmente em ambiente amplamente adotado em soluções de *Big Data*.

O sistema deverá se comunicar com o banco de dados *NoSQL MongoDB* para armazenar o conteúdo requerido pelo usuário.

O *software* deverá interagir com a *search engine Elasticsearch* para permitir *full-text search*.

O *software* deverá possuir *Graphical User Interface (GUI)* simples.

O Diário Oficial da União deverá ser disponibilizado em portal *web* periodicamente.

### 3.2 Caso de Uso

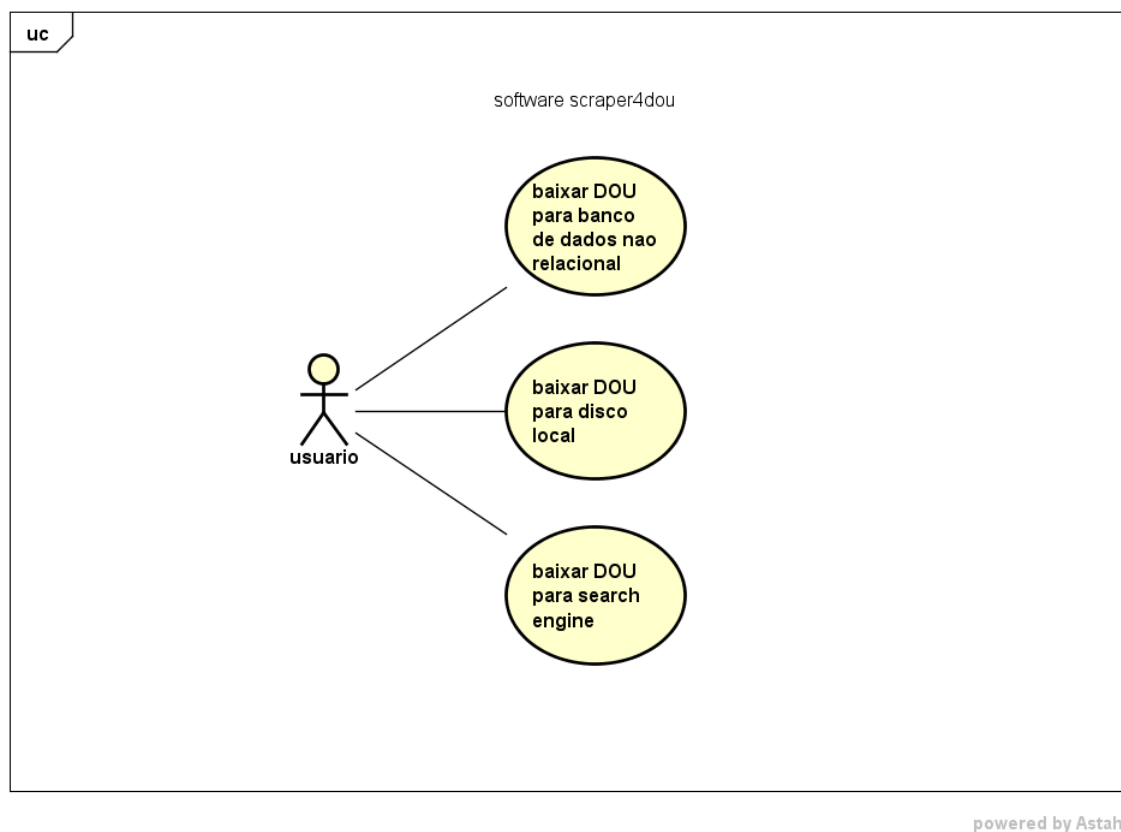
#### 3.2.1 Visão geral do Caso de Uso

O diagrama de Caso de Uso abaixo descreve o relacionamento entre o ator e as funcionalidades do *software*. No cenário descrito existe somente um ator, o usuário, que interage com o *software*.

### 3.2.2 Diagrama de Caso de Uso

Descreve uma sequência de ações que resulta em algo mensurável para um ator, que pode ser uma pessoa, organização ou sistema externo que faz interação com o *software* (AMBLER, 2004, p. 134).

Figura 1: Diagrama de Caso de Uso



Fonte: Pesquisa de campo (2016)

### 3.2.3 Especificação dos Casos de Uso

#### 3.2.3.1 Caso de Uso Baixar DOU para disco local

Tabela 8: Caso de Uso - Baixar DOU para disco local

CASO DE USO	
Baixar DOU para disco local	
Este Caso de Uso permite realizar o <i>download</i> do DOU para o disco local	
ATORES	
NOME DO ATOR	DESCRIÇÃO
<i>Data Scientist</i>	Responsável por baixar arquivos do DOU
FLUXO PRINCIPAL	
AÇÕES DOS ATORES	AÇÕES DO SISTEMA
1. Indicar data ou intervalo de data	
2. Indicar caderno ou intervalo de caderno	
3. Selecionar "disco local" no <i>ComboBox</i>	4. Permitir selecionar pasta de destino
5. Especificar caminho válido de pasta	
6. Clicar no botão de <i>download</i>	7. Realizar requisições e salvar arquivos
PRÉ-CONDIÇÕES	
Conexão com a <i>Internet</i> ativa	

Fonte: Pesquisa de campo (2016)

#### 3.2.3.2 Caso de Uso Baixar DOU para *NoSQL*

Tabela 9: Caso de Uso - Baixar DOU para *NoSQL*

CASO DE USO	
Baixar DOU para <i>NoSQL</i>	
Este Caso de Uso permite realizar o <i>download</i> do DOU para <i>NoSQL</i>	
ATORES	
NOME DO ATOR	DESCRIÇÃO
<i>Data Scientist</i>	Responsável por baixar arquivos do DOU
FLUXO PRINCIPAL	
AÇÕES DOS ATORES	AÇÕES DO SISTEMA
1. Indicar data ou intervalo de data	
2. Indicar caderno ou intervalo de caderno	
3. Selecionar " <i>MongoDB</i> " no <i>ComboBox</i>	4. Permitir informar <i>URL</i> de destino
5. Especificar <i>URL</i> válida para servidor <i>MongoDB</i>	
6. Clicar no botão de <i>download</i>	7. Realizar requisições e salvar conteúdo
PRÉ-CONDIÇÕES	
Conexão com a <i>Internet</i> ativa	

Fonte: Pesquisa de campo (2016)

3.2.3.3 Caso de Uso Baixar DOU para *search engine*Tabela 10: Caso de Uso - Baixar DOU para *search engine*

CASO DE USO	
Baixar DOU para <i>search engine</i>	
Este Caso de Uso permite disponibilizar o conteúdo do DOU para <i>search engine</i>	
ATORES	
NOME DO ATOR	DESCRIÇÃO
<i>Data Scientist</i>	Responsável por baixar arquivos do DOU
FLUXO PRINCIPAL	
AÇÕES DOS ATORES	AÇÕES DO SISTEMA
1. Indicar data ou intervalo de data	
2. Indicar caderno ou intervalo de caderno	
3. Selecionar " <i>Elasticsearch</i> " no <i>ComboBox</i>	4. Permitir informar <i>URL</i> de destino
5. Especificar <i>URL</i> válida para <i>engine Elasticsearch</i>	
6. Clicar no botão de <i>download</i>	7. Realizar requisições e salvar conteúdo
PRÉ-CONDIÇÕES	
Conexão com a <i>Internet</i> ativa	

Fonte: Pesquisa de campo (2016)

## 3.3 Diagrama de Projeto

## 3.3.1 Especificação das Classes

Especificação das classes.

## 3.4 Diagramas de Sequência

Diagramas de Sequência.

## 3.5 Diagrama de Estado

Diagrama de Estado.

## 3.6 Interface Visual do Sistema

Interface Visual do Sistema.

## 4 Conclusão e trabalhos futuros

### 4.1 Conclusão

Conclusão.

### 4.2 Trabalhos futuros

Esse projeto é parte inicial de um escopo maior, pois trata-se de uma engrenagem na percepção de informações a partir de dados disponíveis no Diário Oficial da União.

Com a possibilidade de interação com os dados a partir de um mecanismo de busca propiciada por esse projeto, encaminha-se o desenvolvimento de solução para pesquisa textual, facilitando implementação futura de solução nesse sentido.

Como importante fonte oficial de comunicação, esse canal será explorado ainda com *softwares* de visualização de dados, tais como *Kibana*, para melhor percepção de números divulgados, valores, nomeações, exonerações, etc.

## Referências

- AMBLER, S. W. *The Object Primer: Agile Model-Driven Development with UML 2.0*. third. [S.l.]: Cambridge University Press, 2004. Citado na página 18.
- BANKER, K. et al. *MongoDB in Action*. second. [S.l.]: Manning Publications Company, 2016. Citado na página 14.
- DB-ENGINE. 2016. Acessado em: 20/10/2016. Disponível em: <<http://db-engines.com/en/about>>. Citado na página 15.
- DRISCOLL, M. *Python 101*. [S.l.]: You Lulu Incorporated, 2014. Citado na página 13.
- GHEORGHE, R.; HINMAN, M.; RUSSO, R. *Elasticsearch in Action*. [S.l.]: Manning Publications Company, 2015. Citado na página 15.
- MARTELLI, A. *Python in a Nutshell: A Desktop Quick Reference*. second. [S.l.]: O'Reilly Media, 2006. Citado na página 13.
- PSF. 2016. Acessado em: 20/10/2016. Disponível em: <<https://docs.python.org/3/faq/general.html>>. Citado na página 13.
- ROBBINS, A.; HANNAH, E.; LAMB, L. *Learning the vi and Vim Editors: Text Processing at Maximum Speed and Power*. seventh. [S.l.]: O'Reilly Media, 2008. Citado na página 14.