



# FERRAMENTA PARA TRATAMENTO DE INFORMAÇÕES SENSÍVEIS EM BANCOS DE DADOS

Varlen Pavani Neto

Projeto de Graduação apresentado ao Curso de Engenharia Eletrônica e de Computação da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Engenheiro.

Orientador: Heraldo Luis Silveira de Almeida

Rio de Janeiro  
Fevereiro de 2020

# FERRAMENTA PARA TRATAMENTO DE INFORMAÇÕES SENSÍVEIS EM BANCOS DE DADOS

Varlen Pavani Neto

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO  
DE ENGENHARIA ELETRÔNICA E DE COMPUTAÇÃO DA ESCOLA PO-  
LITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO  
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU  
DE ENGENHEIRO ELETRÔNICO E DE COMPUTAÇÃO

Autor:

---

Varlen Pavani Neto

Orientador:

---

Prof. Heraldo Luis Silveira de Almeida

Examinador:

---

Prof xxxxx

Examinador:

---

Prof xxxxx

Rio de Janeiro

Julho de 2020

## Declaração de Autoria e de Direitos

Eu, *Varlen Pavani Neto* CPF 142.722.117-06, autor da monografia *FERRAMENTA PARA TRATAMENTO DE INFORMAÇÕES SENSÍVEIS EM BANCOS DE DADOS*, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na disciplina de Projeto de Graduação da Escola Politécnica da UFRJ é de sua autoria, sendo original em forma e conteúdo.
2. Excetua-se do item 1. eventuais transcrições de texto, figuras, tabelas, conceitos e idéias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.
3. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.
4. O autor pode, excepcionalmente, encaminhar à Comissão de Projeto de Graduação, a não divulgação do material, por um prazo máximo de 01 (um) ano, improrrogável, a contar da data de defesa, desde que o pedido seja justificado, e solicitado antecipadamente, por escrito, à Congregação da Escola Politécnica.
5. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.
6. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.
7. Por ser verdade, firmo a presente declaração.

---

Varlen Pavani Neto

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es).

## DEDICATÓRIA

Opcional.

## **AGRADECIMENTO**

Sempre haverá. Se não estiver inspirado, aqui está uma sugestão: dedico este trabalho ao povo brasileiro que contribuiu de forma significativa à minha formação e estada nesta Universidade. Este projeto é uma pequena forma de retribuir o investimento e confiança em mim depositados.

## **RESUMO**

Inserir o resumo do seu trabalho aqui. O objetivo é apresentar ao pretense leitor do seu Projeto Final uma descrição genérica do seu trabalho. Você também deve tentar despertar no leitor o interesse pelo conteúdo deste documento.

Palavras-Chave: trabalho, resumo, interesse, projeto final.

## ABSTRACT

Insert your abstract here. Insert your abstract here. Insert your abstract here.  
Insert your abstract here. Insert your abstract here.

Key-words: word, word, word.



## SIGLAS

UFRJ - Universidade Federal do Rio de Janeiro

WYSIWYG - *What you see is what you get*

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Tema . . . . .	1
1.2	Delimitação . . . . .	2
1.3	Justificativa . . . . .	2
1.4	Objetivos . . . . .	2
1.5	Metodologia . . . . .	2
1.6	Descrição . . . . .	3
<b>2</b>	<b>Privacidade na Era da Informação</b>	<b>4</b>
2.1	O entendimento multifacetado de privacidade . . . . .	4
2.2	A General Data Protection Regulation europeia . . . . .	4
2.3	A Lei Geral de Protecao de Dados brasileira . . . . .	4
<b>3</b>	<b>Metodologia</b>	<b>5</b>
3.1	Descoberta de Ferramentas . . . . .	5
3.1.1	Anonymizer . . . . .	5
3.1.2	Data::Anonymization . . . . .	6
3.1.3	ARX - Open Source Data Anonymization Software . . . . .	6
3.1.4	Presidio . . . . .	6
	<b>Bibliografia</b>	<b>7</b>
<b>A</b>	<b>O que é um apêndice</b>	<b>8</b>
<b>B</b>	<b>Encadernação do Projeto de Graduação</b>	<b>10</b>
<b>C</b>	<b>O que é um anexo</b>	<b>12</b>

# Lista de Figuras

B.1	Encadernação do projeto de graduação. . . . .	11
-----	---	----

# Lista de Tabelas

# Capítulo 1

## Introdução

O presente capítulo tem como objetivo apresentar brevemente o escopo do trabalho desenvolvido assim como sua motivação, enumerando conceitos pertinentes a área de conhecimento.

### 1.1 Tema

Este projeto tem como tema o estudo de técnicas para anonimização e dessensibilização em conjuntos de dados. Especificamente, serão avaliadas diferentes técnicas e suas implementações em software de código aberto.

Este é um trabalho majoritariamente de Engenharia de Software, contemplando um ciclo de vida para planejamento, pesquisa e implementação da solução proposta.

É possível dizer que se trata de um trabalho da Área de Engenharia de Dados, dada a natureza das entidades que serão manipuladas. Também serão vistos conceitos da área de Segurança Digital, especificamente Direito Digital e Privacidade.

Assim, este projeto implementa uma nova ferramenta de anonimização e dessensibilização de código aberto englobando mais funcionalidades do que outras ferramentas existentes atualmente.

## 1.2 Delimitação

Este trabalho se limita a estudar técnicas de anonimização, garantia de privacidade e dessensibilização, criadas até o presente momento de sua concepção, que permitam compatibilizar um banco de dados contendo informações sensíveis a uso por terceiros. As implementações destas técnicas, quando existentes, serão estudadas a partir de softwares de código aberto.

## 1.3 Justificativa

TODO

Apresentar o porquê do tema ser interessante de ser estudado. Cuidado, não é a motivação particular. Devem ser apresentadas razões para que alguém deva se interessar no assunto, e não quais foram suas razões particulares que motivaram você a estudá-lo (tamanho do texto: livre).

## 1.4 Objetivos

O objetivo geral deste trabalho é implementar um software livre que permita processar um conjunto de dados tornando-o dessensibilizado, sendo desta forma possível utilizá-lo para apoiar atividades de desenvolvimento e testes de sistemas sem a preocupação com vazamento de dados por parte de terceiros. Especificamente, o software deve: (1) Remover informações que permitam associar indivíduos com um conjunto de dados; (2) Permitir a substituição de dados reais sensíveis por dados gerados a partir de estatísticas; (3) Respeitar e manter a estrutura do modelo de dados existente, alterando somente o seu conteúdo.

## 1.5 Metodologia

TODO Como é a abordagem do assunto. Como foi feita a pesquisa, se houve validação, etc. Em resumo, você deve explicar qual foi sua estratégia para atender ao objetivo do trabalho (tamanho do texto: livre).

## 1.6 Descrição

No capítulo 2 será feita a apresentação introdutória sobre privacidade, conceitos associados e suas relevâncias no contexto socioeconômico contemporâneo.

O capítulo 3 apresenta um panorama de ferramentas de anonimização cuja licença garanta a abertura do código-fonte com o intuito de prospectar as funcionalidades fornecidas por esses programas.

## Capítulo 2

# Privacidade na Era da Informação

O presente capítulo propõe-se a introduzir o leitor ao entendimento acadêmico existente sobre o conceito de privacidade, especificamente no que contempla a área de Tecnologia da Informação com foco nos pontos de interesse ao presente trabalho.

### 2.1 O entendimento multifacetado de privacidade

### 2.2 A General Data Protection Regulation europeia

### 2.3 A Lei Geral de Proteção de Dados brasileira



# Capítulo 3

## Metodologia

A etapa inicial deste trabalho consistiu na prospecção das ferramentas de código aberto existentes com o intuito de entender as capacidades fornecidas e determinar os requisitos necessários a uma nova solução.

### 3.1 Descoberta de Ferramentas

Para descoberta de ferramentas de anonimização neste trabalho, foram utilizados projetos de código-aberto disponíveis em repositórios públicos no GitHub.

Especificamente, foram consideradas as métricas de stars e forks dos repositórios como medida de popularidade para seleção de 4 projetos relativos a anonimização de dados. Todos os projetos apresentavam novas modificações no código-fonte em Março de 2020, indicando que são mantidos pela comunidade. Buscou-se inferir os casos de uso, funcionalidades, usabilidade e tecnologias utilizadas a partir da documentação dos projetos. Considerou-se a primeira linha dos arquivos README dos repositórios como nome do projeto.

#### 3.1.1 Anonymizer

O código-fonte deste projeto está disponível em [github.com/DivanteLtd/anonymizer](https://github.com/DivanteLtd/anonymizer).

Anonymizer foi desenvolvida na linguagem de programação Ruby pela companhia europeia Divante e opera exclusivamente em bancos de dados MySQL. Segundo a documentação, sua funcionalidade mais importante é a formatação de dados. A ferramenta substitui os dados originais por dados gerados.

Sua instalação é feita a partir de uma cópia do repositório de código-fonte para a máquina na qual se deseja executar.

O processo de anonimização é feito a partir de um arquivo de dump do banco de dados. Este arquivo pode estar na mesma máquina que executa o processo ou em uma máquina remota. O usuário deve criar um arquivo no formato JSON com os parâmetros do processo.

A ferramenta permite substituir os valores nas tabelas por valores fixos, valores em branco ou gerados, de acordo com a categoria. São suportadas as categorias `firstname`, `lastname`, `login`, `email`, `telephone`, `company`, `street`, `postcode`, `city`, `full_address`, `vat_id`, `ip`, `quote`, `website`, `iban`, `json`, `uniq_email`, `uniq_login`, `regon` (equivalente polonês ao CNPJ) e `pesel` (equivalente polonês ao CPF). Também é possível truncar todos os dados de uma tabela.

Também é provida a capacidade de executar um comando SQL arbitrário antes ou depois do processo.

### **3.1.2 Data::Anonymization**

O código-fonte deste projeto está disponível em [github.com/sunitparekh/data-anonymization/](https://github.com/sunitparekh/data-anonymization/).

TODO

### **3.1.3 ARX - Open Source Data Anonymization Software**

O código-fonte deste projeto está disponível em [github.com/arx-deidentifier/arx](https://github.com/arx-deidentifier/arx).

TODO

### **3.1.4 Presidio**

O código-fonte deste projeto está disponível em [github.com/microsoft/presidio](https://github.com/microsoft/presidio).

TODO

## Referências Bibliográficas

# Apêndice A

## O que é um apêndice

Elemento que consiste em um texto ou documento elaborado pelo autor, com o intuito de complementar sua argumentação, sem prejuízo do trabalho. São identificados por letras maiúsculas consecutivas e pelos respectivos títulos.

## Apêndice B

# Encadernação do Projeto de Graduação

Número	<p><b>UNIVERSIDADE FEDERAL DO RIO DE JANEIRO</b></p> <p><b>Escola Politécnica</b></p> <p><b>Departamento de Eletrônica e de Computação</b></p>
Nome do Aluno	
Título do Projeto*	<p><b>Título do Projeto</b></p> <p><b>Nome do Aluno</b></p>
Ano	<p><b>Projeto de Graduação</b></p> <p><b>Mês / Ano</b></p>

**\* Título resumido caso necessário**  
**Capa na cor preta, inscrições em dourado**

Figura B.1: Encadernação do projeto de graduação.

## Apêndice C

### O que é um anexo



Documentação não elaborada pelo autor, ou elaborada pelo autor mas constituindo parte de outro projeto.