UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA ELETRÔNICA E DE COMPUTAÇÃO

PROPOSTA DE PROJETO DE GRADUAÇÃO

Aluno: Varlen Pavani Neto
varlenneto@poli.ufrj.br

Orientador: Heraldo Luis Silveira de Almeida

1. TÍTULO

Ferramenta para Tratamento de Informações Sensíveis em Bancos de Dados

2. ÊNFASE

Computação

3. TEMA

O tema do trabalho é o desenvolvimento de uma ferramenta que permita exportar informações dessensibilizadas e anonimizadas de um banco de dados.

4. DELIMITAÇÃO

O objeto de estudo do trabalho são processos de anonimização e garantia de privacidade em conjuntos de dados, culminando na implementação de um software para realizar tais processos. No presente trabalho, as implementações destes processos serão estudados a partir de software livre.

5. JUSTIFICATIVA

A ubiquidade de dados pessoais produzidos pelo consumo de serviços digitais e o seu posterior entendimento como *commodity* [1] a uma crescente preocupação com a privacidade por parte das pessoas.

Por muitas vezes a privacidade é vista como um direito fundamental[2]. Assim, a nível governamental esta preocupação foi transformada em legislação ao redor do mundo, como por exemplo o Regulamento Geral sobre a Proteção de Dados 2016/679 da União Europeia e a brasileira Lei Geral de Proteção de Dados Pessoais, Lei nº 13.709/2018.

Surge então nas instituições a necessidade de adequar processos e a maneira com que tratam seus dados à legislação vigente.

O presente projeto busca então complementar os modelos e técnicas, e revisar ferramentas existentes com o intuito de verificar seus pontos de melhoria e implementar um software de código aberto que auxilie as instituições a atingirem seus objetivos de conformidade com a lei, permitindo processar massas de dados sem perda de utilidade prática.

6. OBJETIVO

O objetivo geral deste trabalho é implementar um software livre que permita processar um banco de dados de modo a torná-lo dessensibilizado, podendo assim ser utilizado para apoiar atividades de desenvolvimento de sistemas sem a preocupação com vazamento de dados por parte de terceiros.

Sendo assim, os objetivos específicos a serem atingidos na implementação são: (1) O software deve remover informações que permitam associar indivíduos a um conjunto de dados; (2) O software deve substituir dados reais sensíveis por dados gerados a partir de estatísticas; (3) O software não deve alterar a estrutura do modelo de dados existente, somente o seu conteúdo.;

7. METODOLOGIA

Este trabalho inicialmente consiste num estudo das técnicas de anonimização de dados anteriormente aplicadas em dados médicos, que possuem necessidade de cuidados especiais de privacidade anterior a existência das leis de proteção de dados gerais, com o intuito de verificar como e quais técnicas podem ser utilizadas fora do contexto da informática para medicina. Para esta etapa inicial será feita uma revisão da literatura do assunto.

Em seguida, serão analisadas soluções de software livre existentes para verificar a implementação das técnicas exploradas anteriormente. Esta etapa contemplará softwares cuja licença seja aberta e estejam disponíveis no Github, tendo por objetivo caracterizar as funcionalidades existentes de modo a definir os requisitos do software a ser implementado, de maneira que este traga vantagens em relação as soluções existentes.

A partir a definição dos requisitos na etapa anterior, tem-se por inicio o de-

senvolvimento do código-fonte da nova solução. Esta solução será implementada utilizando a linguagem de programação Python, por se tratar de uma linguagem aberta e possuir suporte multiplataforma. Também será utilizada a biblioteca SQLAlchemy para implementar as atividades de acesso e manipulação dos bancos de dados, pois esta biblioteca fornece uma camada de abstração sobre diferentes implementações de APIs SQL[3], possibilitando que o projeto possa ser utilizado com diferentes implementações de bancos de dados SQL.

O alvo final deste trabalho é completar a implementação de um Minimum Viable Product com as funcionalidades determinadas anteriormente a ser validado sobre bases de dados públicas disponíveis na internet.

8. MATERIAIS

Serão utilizados software livres para estudo. Linguagem Python. Plataforma Github para controle de versão do código-fonte.

9. CRONOGRAMA

Apresentada graficamente conforme a Figura 1.

Fase 1: Estudo inicial sobre impacto socioeconômico causado pela visão de dados como commoditie

Fase 2: Análise bibliográfica de técnicas de privacidade e tópicos afins.

Fase 3: Estudo e levantamento das características de softwares de anonimização de código-aberto.

Fase 4: Implementação do código-fonte.

Fase 5: Testes com bases de dados públicas.

Fase 6: Revisão da Monografia

Fase 7: Escrita da monografia (Parcialmente concorrente com as demais atividades)

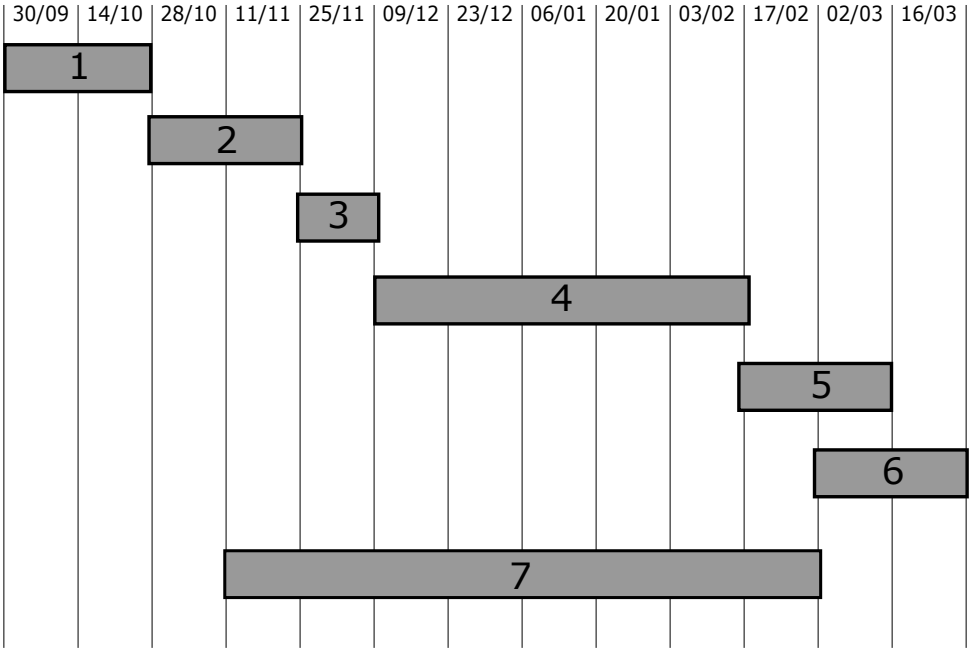

Figura 1: *Planejamento*

Referências Bibliográficas

[1] MEYER, Julie, "Personal data: time to argue about money, not privacy". Financial Times. https://www.ft.com/content/ba39f1a6-ea66-11e4-96ec-00144feab7de#ixzz3YcJONvSE, 2013 (Acesso em 22 setembro 2019)

[2] SOLOVE, Daniel J. "Understanding privacy."Cambridge, MA: Harvard university press, 2008

[3] "SQL Alchemy Features- SQLAlchemy authors and contributors. https://www.sqlalchemy.org/features.html, 2019 (Acesso em 22 setembro 2019)

Rio de Janeiro, 23 de setembro de 2019

Varlen Pavani Neto - Aluno

Heraldo Luis Silveira de Almeida - Orientador