

Desenvolvimento de um Sistema de Recomendação de Candidatos à Cargos Públicos Utilizando Técnicas de Aprendizagem de Máquina e Filtragem de Informação

Projeto 14786

Orientador: Prof. Dr. Fábio Santos da Silva (fssilva@uea.edu.br)

Bolsista: Lucas Pereira Reis (lprs.eng@uea.edu.br)

Engenharia de Computação
Universidade do Estado do Amazonas
Escola Superior de Tecnologia

19 de abril de 2019

Sumário

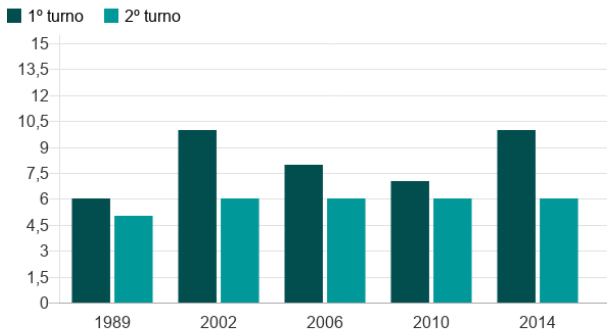
- 1 Introdução
- 2 Justificativa
- 3 Objetivos
- 4 Metodologia
- 5 Fundamentação Teórica
- 6 Desenvolvimento
- 7 Resultados Parciais
- 8 Cronograma
- 9 Referências



Figura 1: Eleições no Brasil. (<https://tinyurl.com/y6knph9d>)

Brancos e Nulos nas Eleições Presidenciais

em % de votos



Fonte: TSE



Figura 2: Gráfico de brancos e nulos nas eleições passadas. (<https://tinyurl.com/y4f6pawl>)

■ %



5 / 36

Sobrecarga de Informação

- Ocorre quando somos submetidos em excesso à mídia, tecnologia e informação. Dificulta à absorção e reflexão de informações, podendo levar à decisões erradas e erros. (<https://tinyurl.com/y3y9toon>)
- A exagerada quantidade de candidatos dificulta a escolha do eleitor, deixando-o perdido durante as eleições.
- Já existe uma solução para ajudar na escolha?



Figura 4: Trabalhos relacionados à recomendação de candidatos à cargos públicos.

Objetivo

- Desenvolver um sistema de recomendação de candidatos à órgãos públicos para auxiliar os eleitores durante a escolha.
- A recomendação irá mensurar o grau de relevância dos candidatos à determinados temas correlacionando com as avaliações feitas pelo eleitor/usuário em relação aos temas e candidatos.

Objetivos Específicos

- Realizar revisão bibliográfica (1).
- Investigar e pesquisar tecnologias apropriadas para o desenvolvimento do projeto; (2)
- Realizar a aquisição de dados sobre políticos através da internet; (3)
- Realizar pré-processamento dos dados, gerando o *dataset*; (4)
- Projetar a arquitetura do sistema e seus componentes; (5)
- Desenvolver um protótipo para web integrando-o com o sistema gerado; (6)
- Realizar testes para avaliar a qualidade das recomendações geradas pelo sistema. (7)

Tabela 1: Metodologia utilizada no desenvolvimento do projeto.

Objetivos	Metodologia
1	Estudo sobre sistemas de recomendação, filtragem de informação e aprendizagem de máquina.
2	Pesquisar sobre linguagens de programação, tipos de técnica de filtragem e tipos de arquitetura <i>web</i> que serão utilizadas.
3	Realizar a coleta ou a busca de dados relacionadas ao candidatos e usuários utilizando as técnicas definidas em 2.
4	Realizar a limpeza e pré-processamento dos dados obtidos para então definir os <i>datasets</i> .
5	Definir como o sistema será estruturado para implementar as técnicas estipuladas em conjunto com os <i>datasets</i> .
6	Desenvolver uma interface para o sistema com as ferramentas definidas em 2.
7	Utilizar métricas de desempenho para avaliar a recomendação gerada.



Figura 5: Ferramentas utilizadas para o desenvolvimento do projeto até agora: Python, Git/Github, Flask e SQLite.

Filtragem Colaborativa

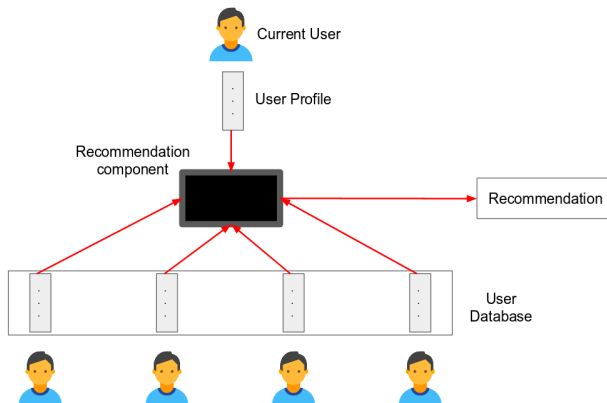


Figura 6: Exemplo do processo de filtragem colaborativa. (Autor)

Filtragem Baseada em Conteúdo

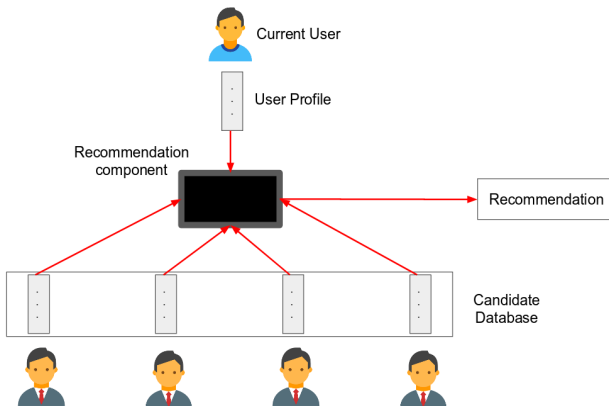


Figura 7: Exemplo do processo de filtragem baseada em conteúdo. (Autor)

Filtragem Híbrida

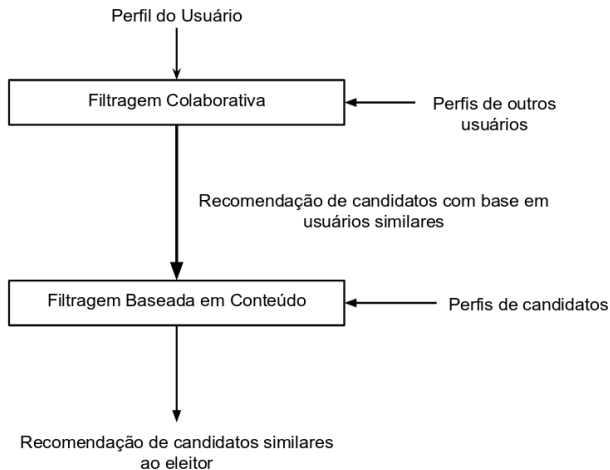


Figura 8: Fluxograma da implementação da Filtragem Híbrida. (Autor)

Métricas de similaridade

São utilizadas para calcular o quão próximo o usuário é de outro usuário e vice-versa.

- Métricas mais utilizadas:
 - Distância de *Manhattan*.
 - Distância Euclidiana.
 - Correlação de *Pearson*.
 - Similaridade Cosseno.
- Em um ambiente 2D, cada pessoa é representada por um ponto (x,y) .

Distância de *Manhattan*

- É o valor absoluto da diferença entre os valores x mais o valor absoluto da diferença entre os valores de y .
- Utilizada para dados densos sem grau de inflação.

$$d(x, y) = \left(\sum_{k=1}^n |x_k - y_k|^r \right)^{\frac{1}{r}}$$

- $r = 1$: distância de *Manhattan*.
- $r = 2$: distância Euclidiana.

TF - IDF (*Term Frequency - Inverse Document Frequency*)

- Técnica de recuperação de informação utilizada para mensurar a frequência do termo (TF) e sua frequência inversa no documento (IDF).
- Cada palavra possui seu próprio TF e IDF, o produto das técnicas é o peso TF-IDF do termo. (<http://tinyurl.com/yxkmlmt>)

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

- $tf_{i,j}$: número de ocorrências do termo i em j .
- df_i : número de documentos contendo i .
- N : número total de documentos.

$$\text{Precisão} = \frac{\text{Candidatos relevantes recomendados}}{\text{Total de candidatos recomendados}}$$

$$\text{Revocação} = \frac{\text{Candidatos relevantes recomendados}}{\text{Total de candidatos relevantes}}$$

$$\text{Medida-F} = \frac{2 \times (\text{Precisão} \times \text{Revocação})}{\text{Precisão} + \text{Revocação}}$$

Extração de Dados

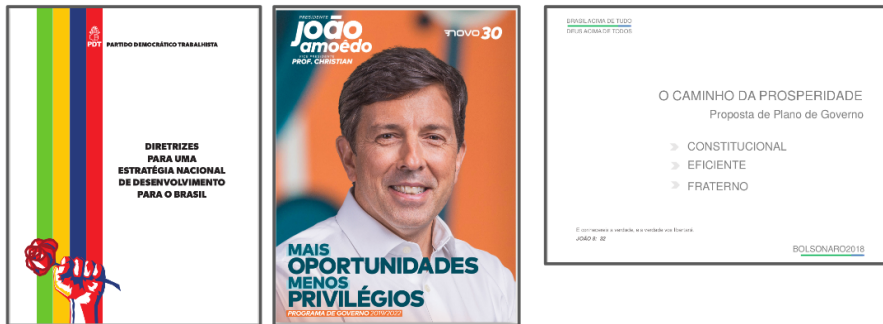


Figura 9: A extração de dados dos candidatos foram obtidas a partir de suas propostas. (<http://tinyurl.com/y4ay5vxu>)

Extração de Dados

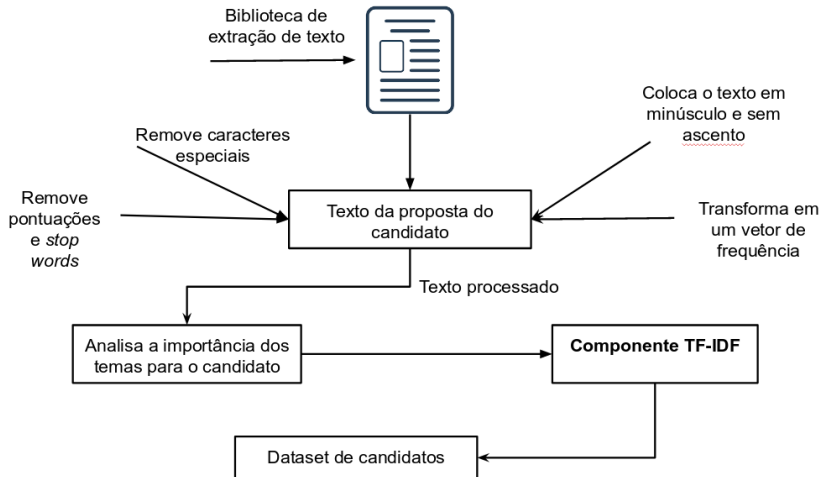
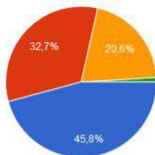


Figura 10: Fluxograma do processo de extração de dados da proposta. (Autor)

Dados de Usuário Reais

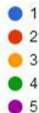
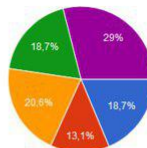
Geraldo Alckmin

107 respostas



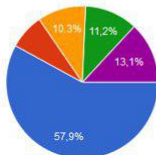
João Amoêdo

107 respostas



Jair Bolsonaro

107 respostas



Guilherme Boulos

107 respostas

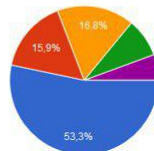
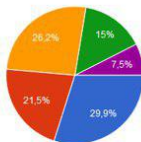


Figura 11: Análise das respostas dos usuários no formulário.

Dados de Usuários Reais

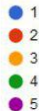
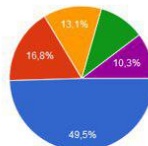
Ciro Gomes

107 respostas



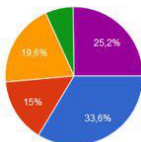
Fernando Haddad

107 respostas



Marina Silva

107 respostas



Cabo Daciolo

107 respostas

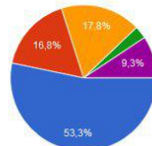


Figura 12: Segunda análise das respostas dos usuários.

Arquitetura do Protótipo Web

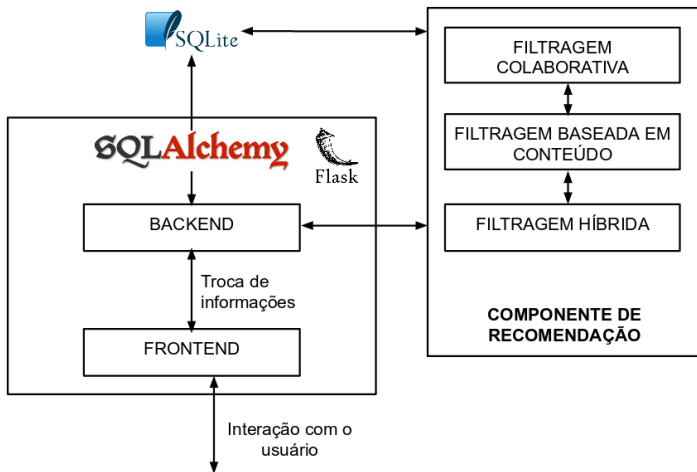


Figura 13: Arquitetura utilizada para o protótipo. (Autor)

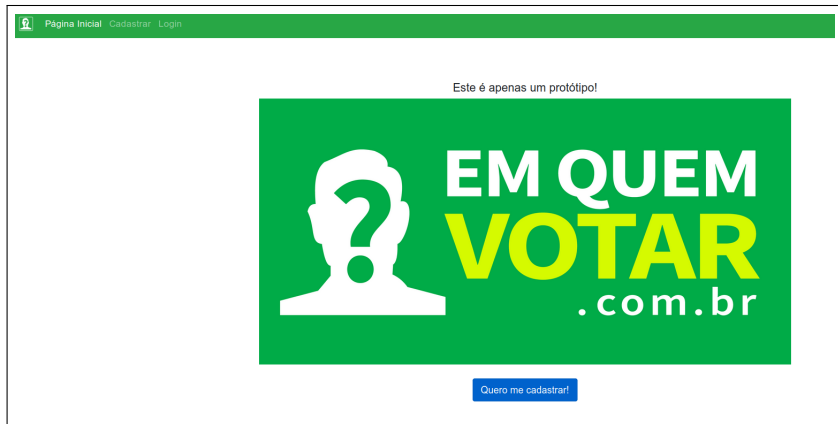



Figura 15: Tela inicial do protótipo.



Por favor insira seus dados

Cadastrar

Usuário


Nome

Senha

Cadastrar

Já possui uma conta? [Faça seu login!](#)

Figura 16: Tela de cadastro do protótipo.



Por favor insira seus dados

Login

Usuário

Senha

Entrar

Novo no sistema? [Cadastre-se agora!](#)

Figura 17: Tela de login do protótipo.

Avalie os temas que você mais tem interesse nas propostas de um candidato.

Cultura

Economia

Educação

Meio Ambiente

Saúde

Segurança

Tecnologia

Avaliar

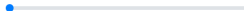
Figura 18: Avaliação de áreas no protótipo.



Avalie os seguintes candidatos conforme sua afinidade com cada um.



Geraldo Alckmin



João Amoêdo

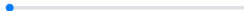



Figura 19: Avaliação de candidatos no protótipo.



Bem Vindo, Altair (**altair**)

Solicitar Recomendação

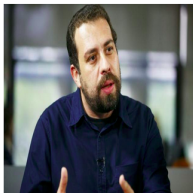
Temas
Cultura 5
Economia 5
Educação 4
Meio Ambiente 3
Saúde 5
Segurança 5
Tecnologia 5
Clique aqui para alterar sua avaliação!

Candidatos
Ciro Gomes 4
Cabo Daciolo 1
Marina Silva 1
Clique aqui para alterar sua avaliação!

Figura 20: Perfil do usuário.



Os seguintes candidatos foram recomendados para você



Guilherme Boulos



Fernando Haddad

Figura 21: Tela de recomendação de candidatos para o usuário.

Tabela 2: Desempenho do sistema por cada usuário,

Usuário	Precisão	Revocação	Medida-F	Nota
1	1.0	0.67	0.8	4.0
2	1.0	0.25	0.4	5.0
3	0.67	0.67	0.67	3.0
4	0.0	0.0	0.0	4.0
5	0.0	0.0	0.0	3.0
6	0.0	0.0	0.0	4.0
7	0.33	1.0	0.5	5.0
8	0.0	0.0	0.0	3.0

Continua

Tabela 3: Desempenho do sistema por cada usuário.

Usuário	Precisão	Revocação	Medida-F	Nota
9	0.5	0.5	0.5	4.0
10	0.33	0.5	0.4	4.0
11	1.0	0.33	0.5	4.0
12	1.0	0.33	0.5	3.0
13	0.67	0.67	0.67	4.0
14	1.0	1.0	1.0	5.0
Média	53.57%	42.26%	42.38%	3.93

Tabela 4: Cronograma de atividades. Verde: concluído, amarelo: em andamento.

Mês	Atividades
08/18	Estudo sobre sistemas de recomendação e trabalhos relacionados
09/18 à 10/18	Estudo de técnicas de aprendizagem de máquina e filtragem de informação
11/18	Identificação dos principais requisitos do sistema, projeto da arquitetura e componentes
12/18 à 02/19	Desenvolvimento de componentes de filtragem de informação e aprendizagem de máquina
03/19 à 04/19	Desenvolvimento do <i>Front-End</i> e <i>Back-End</i>
05/19 à 06/19	Testes e Avaliação do Sistema
Todos	Elaboração de Relatórios e Artigos

- PACIEVITCH, Thais. Eleições no Brasil, Infoescola. Disponível em: <<https://www.infoescola.com/direito/eleicoes-no-brasil/>>.
- BURGARELLI, Rodrigo. Voto inconsciente: o que o brasileiro leva em conta para decidir seu candidato a prefeito?, Estadão. Disponível em: <http://alias.estadao.com.br/noticias/geral,voto-inconsciente-o-que-o-brasileiro-levaem-conta-para-decidir-seu-candidato-para-prefeito,10000077956>.
- ADOMAVICIUS, G.; TUZHILIN, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. In: IEEE Transactions on Knowledge and Data Engineering, v.17, n.6, p.734-749, 2005.

- MCCREA, N. (2015). An Introduction to Machine Learning Theory and Its Applications: A Visual Tutorial with Examples. Acesso em Dezembro de 2017, disponível em :
<<https://www.toptal.com/machine-learning/machine-learning-theoryan-introductoryprimer/>>.
- MALAEB, M. Recall and Precision at k for Recommender Systems. Disponível em: <https://medium.com/@m_n_malaeb/recall-and-precision-at-k-for-recommender-systems-618483226c54/>.
- Lim, K., Kim, C., Kim, G., & Choi, H. (2017, July). A Recommender System for Political Information Filtering. In International Conference of Design, User Experience, and Usability (pp. 129-145). Springer, Cham.

Desenvolvimento de um Sistema de Recomendação de Candidatos à Cargos Públicos Utilizando Técnicas de Aprendizagem de Máquina e Filtragem de Informação

Projeto 14786

Orientador: Prof. Dr. Fábio Santos da Silva (fssilva@uea.edu.br)

Bolsista: Lucas Pereira Reis (lprs.eng@uea.edu.br)

Engenharia de Computação
Universidade do Estado do Amazonas
Escola Superior de Tecnologia

19 de abril de 2019