

WikipediaAnalyzer



Manual do Usuário

Desenvolvido por: Éric Gabriel Kundlatsch

Sumário

1. Introdução	3
2. Requisitos	3
3. Ambiente de Execução	3
4. Instalação de Dependências	3
5. Estrutura Geral	3
6. Módulo I – Análise de Mudança de Vieses ao Longo do Tempo	4
6.1 Coleta de Versões de um Verbetes	4
6.2 Geração de Gráfico de Total de Palavras por Versão	4
6.3 Frequência de Palavras Específicas	4
6.4 Gráfico de TF-IDF	4
6.5 Palavras Mais Frequentes	4
7. Módulo II – Análise de Redes de Usuários	5
7.1 Coleta de Edições	5
7.2 Classificação de Usuários	5
7.3 Detecção de Reversões	5
7.4 Construção e Visualização da Rede	5
8. Módulo III – Integração com Voyant Tools	5
8.1 Geração de ZIP com Versões Limpas	5
8.2 Coleta de Links para Versões Históricas	5
9. Considerações Finais	6

1. Introdução

WikipediaAnalyzer é uma ferramenta interativa desenvolvida com fins acadêmicos para apoiar a análise de verbetes da Wikipédia, buscando compreender mudanças de vieses e narrativas ao longo do tempo. Desenvolvida como parte de um Trabalho de Conclusão de Curso (TCC) no curso de História da Universidade Federal de Santa Catarina (UFSC), essa ferramenta busca oferecer aos historiadores uma forma acessível de visualizar mudanças no conteúdo e nos padrões de edição colaborativa da enciclopédia livre.

A ferramenta foi construída em Python, mas é projetada para ser utilizada exclusivamente através do Google Colab — uma plataforma gratuita e online que permite executar scripts em nuvem, sem necessidade de instalação local. Não é necessário nenhum conhecimento prévio em programação para utilizar o WikipediaAnalyzer, pois o funcionamento é guiado por menus interativos e instruções diretas.

2. Requisitos

- Conta Google (para acesso ao Google Colab);
 - Navegador de internet (Google Chrome, Mozilla Firefox, etc.);
 - Conexão estável com a internet;
-

3. Ambiente de Execução

O código é projetado para rodar diretamente no **Google Colab**, uma ferramenta gratuita oferecida pela Google para execução de código em Python na nuvem. Nenhuma instalação é necessária além de acessar o link do Colab com o código do WikipediaAnalyzer (fornecido pelo autor ou disponibilizado em repositório).

4. Instalação de Dependências

As bibliotecas necessárias são instaladas automaticamente em células no próprio Colab. O usuário apenas precisa executar essas células clicando no botão "play" ao lado de cada bloco de código. O sistema informará em caso de sucesso ou erro na instalação.

5. Estrutura Geral

A ferramenta é dividida em três grandes módulos:

- **Módulo I:** Coleta de versões de um verbete e análise textual (palavras, frequência, variação de vocabulário ao longo do tempo);
- **Módulo II:** Análise da colaboração entre usuários que editaram um verbete específico (com detecção de conflitos e redes de interação);
- **Módulo III:** Exportação das versões do verbete para uso em ferramentas externas, como o Voyant Tools.

Cada módulo é independente e pode ser executado separadamente, de acordo com o objetivo da pesquisa.

6. Módulo I – Análise de Mudança de Vieses ao Longo do Tempo

Esse módulo permite baixar todas as versões de um verbete da Wikipédia e analisá-las com foco na linguagem utilizada. As análises envolvem desde contagem de palavras até gráficos que mostram como determinadas palavras ganham ou perdem destaque ao longo do tempo.

6.1 Coleta de Versões de um Verbetes

O sistema solicita que o usuário digite o nome do verbete da Wikipédia. Em seguida, todas as versões disponíveis desse verbete são baixadas automaticamente e salvas em arquivos de texto. Essas versões formam a base para as análises seguintes.

6.2 Geração de Gráfico de Total de Palavras por Versão

Essa função permite observar o crescimento ou redução do conteúdo textual do verbete ao longo do tempo, medindo quantas palavras existem em cada versão.

6.3 Frequência de Palavras Específicas

O sistema solicita que o usuário digite palavras que deseja acompanhar (por exemplo: "democracia, golpe"). Ele então mostra um gráfico com a frequência de cada uma dessas palavras em todas as versões salvas.

6.4 Gráfico de TF-IDF

Essa função mede a "importância relativa" das palavras ao longo do tempo. Mesmo palavras pouco frequentes podem ter destaque se forem únicas em determinada versão. O usuário insere as palavras que deseja observar e o gráfico mostra sua pontuação TF-IDF ao longo das edições.

6.5 Palavras Mais Frequentes

Essa opção mostra quais são as palavras mais utilizadas em todo o histórico do verbete. O sistema ignora palavras comuns (como "de", "e", "a") e destaca os termos mais relevantes.

7. Módulo II – Análise de Redes de Usuários

Esse módulo permite observar a dinâmica entre os editores da Wikipédia. Ao fornecer um verbete, o sistema constrói uma rede de quem editou, quem reverteu quem e como se dá a colaboração.

7.1 Coleta de Edições

O sistema acessa a Wikipédia e baixa o histórico completo de edições do verbete, incluindo nome de usuário, data, comentários e marcas de reversão.

7.2 Classificação de Usuários

Os usuários são automaticamente classificados como:

- Administradores
- Bots (robôs)
- Autorrevisores
- Editores comuns

Essa classificação é baseada em informações públicas nas páginas de usuário da Wikipédia.

7.3 Detecção de Reversões

Além de editar, usuários podem desfazer ou reverter edições alheias. O sistema identifica esses conflitos e registra quem reverteu quem.

7.4 Construção e Visualização da Rede

Com base nos dados coletados, é construída uma rede visual. Cada ponto representa um usuário. Linhas azuis indicam colaborações (edições próximas), e linhas vermelhas indicam conflitos (reversões). O tamanho dos pontos indica quantas edições cada usuário fez.

8. Módulo III – Integração com Voyant Tools

Esse módulo prepara os dados para serem utilizados em outras ferramentas digitais de análise textual, como o Voyant Tools (<https://voyant-tools.org/>).

8.1 Geração de ZIP com Versões Limpas

O sistema baixa as versões do verbete, remove marcas de formatação da Wikipédia (como links, cabeçalhos, códigos etc.) e gera um arquivo ZIP com os textos em formato limpo. Esse arquivo pode ser carregado diretamente no Voyant Tools para novas análises.

8.2 Coleta de Links para Versões Históricas

Além dos textos, o sistema também pode listar os links diretos para cada versão do verbete, com data e horário de publicação. Isso permite cruzamento de dados com outras fontes.

9. Considerações Finais

WikipediaAnalyzer foi projetado com o objetivo de democratizar o acesso a métodos computacionais na pesquisa histórica. Com ele, é possível explorar a história da própria Wikipédia, seus debates editoriais, mudanças de linguagem e dinâmicas de poder nos bastidores da construção do conhecimento.

Apesar de utilizar linguagens e bibliotecas de programação, o sistema foi totalmente adaptado para funcionar de forma amigável em ambiente Google Colab, sem exigir conhecimentos técnicos.

O uso da Wikipédia como fonte histórica exige crítica e rigor, e ferramentas como essa podem auxiliar o historiador a observar não apenas "o que está escrito", mas "como foi escrito" e "quem escreveu".
