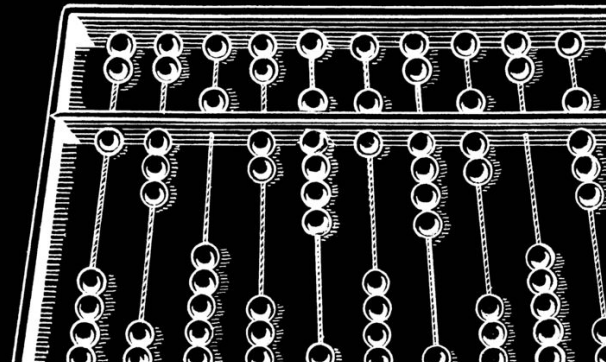


Um Modelo de Classificação de Documentação para Novatos em Software Livre

Candidato: Luiz Felipe Fronchetti Dias
Orientador: Prof. Dr. Marco Aurélio Gerosa

Dissertação apresentada ao
Instituto de Matemática e Estatística
da Universidade de São Paulo para obtenção do título de
Mestre em Ciência da Computação



Introdução

- “Software livre” → software que respeita a liberdade e o senso de comunidade dos usuários;
- Liberdade de ler, estudar, modificar e redistribuir o software desenvolvido.

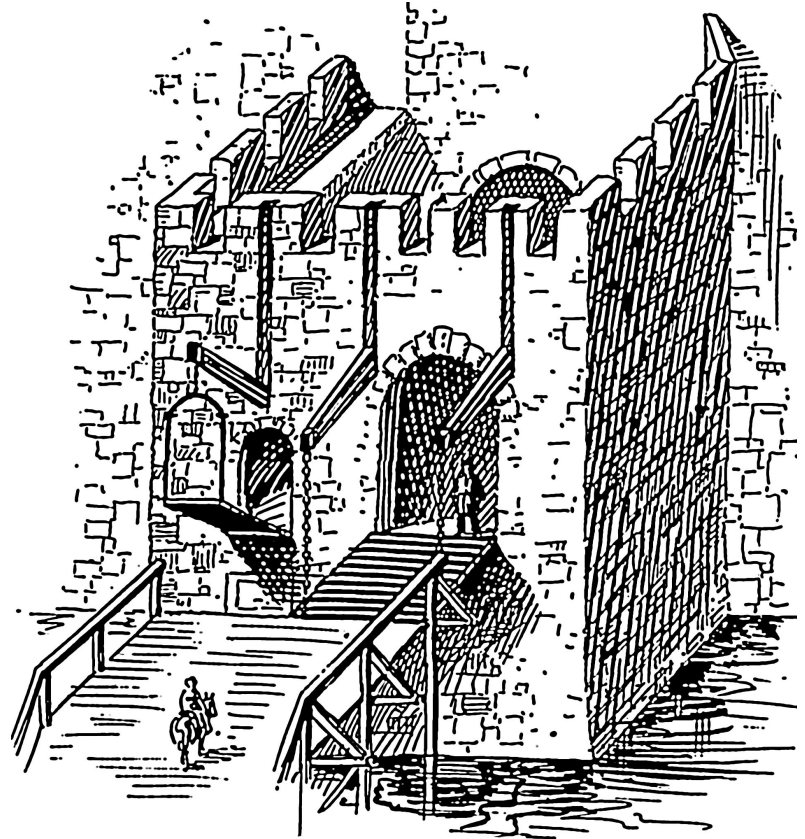


Imagem: Wikimedia, Domínio Público

Introdução

- Tradicionalmente, comunidades que gerem projetos de software livre são compostas e mantidas pela contribuição de voluntários geograficamente distribuídos ao redor do mundo.
- As contribuições são feitas de maneira remota via internet.

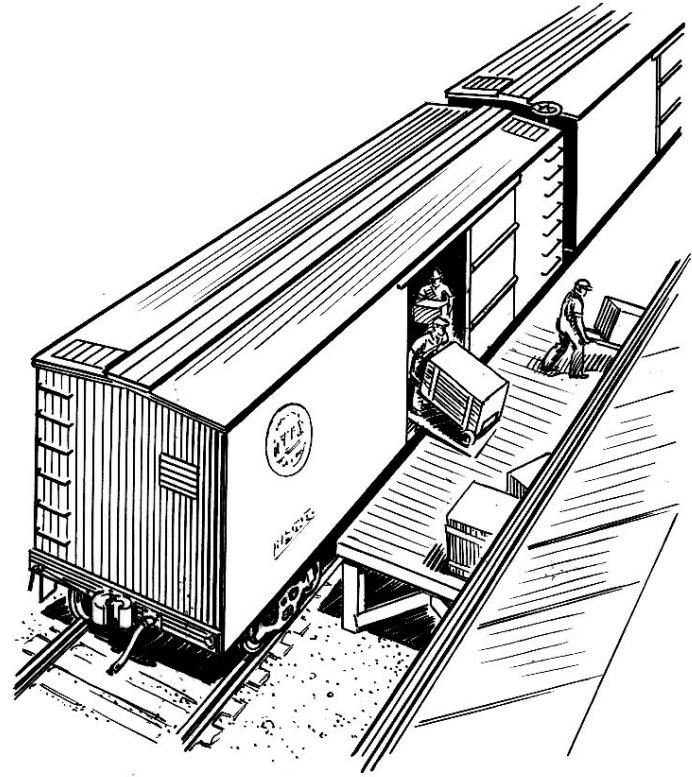


Imagem: Wikimedia, Domínio Público

Problema

- Novos contribuidores costumam enfrentar barreiras ao tentar ingressar em projetos de software livre;
- O acúmulo de barreiras acaba por atrapalhar o processo de submissão, acarretando na desistência de novatos.

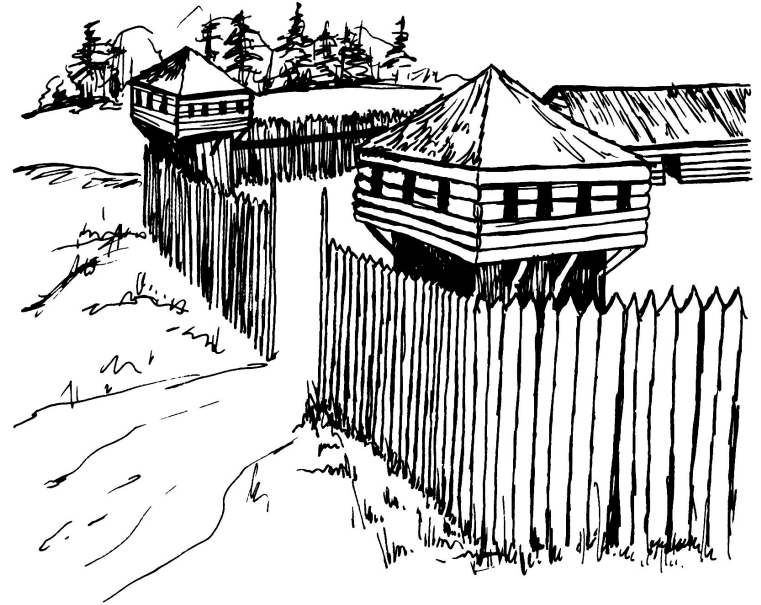


Imagem: Wikimedia, Domínio Público

Problema

- Entre as barreiras, encontram-se problemas associados à documentação dos projetos;
- Falta de clareza e sobrecarga de informações são exemplos de problemas de documentação.

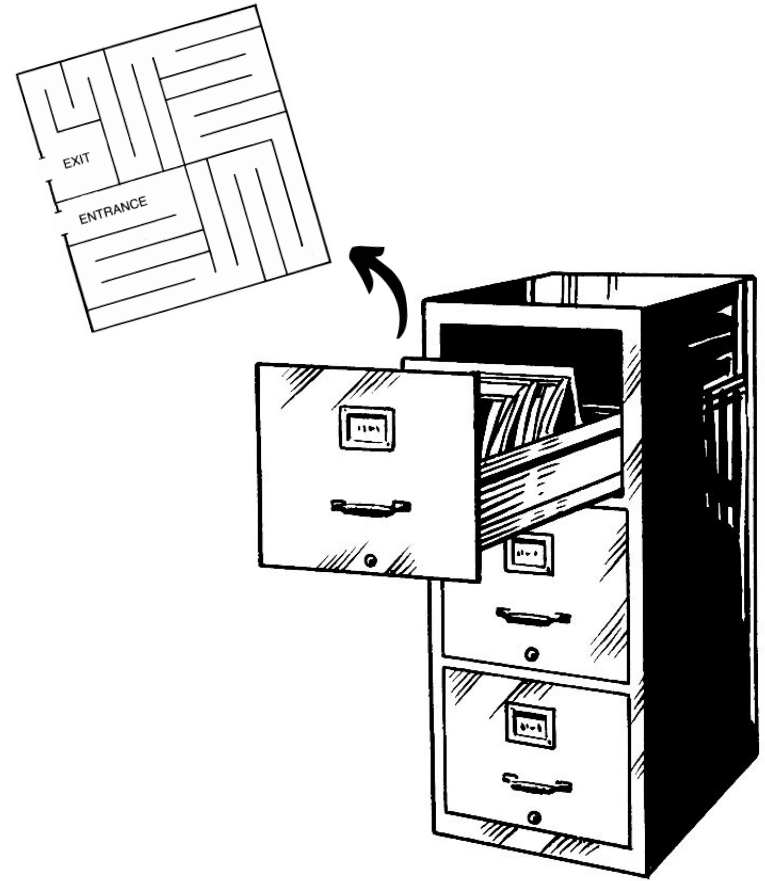


Imagem: Wikimedia, Domínio Público

Proposta

- Criação de um modelo de classificação que identifique trechos em arquivos de contribuição de projetos de software livre que sejam relevantes a novos contribuidores.

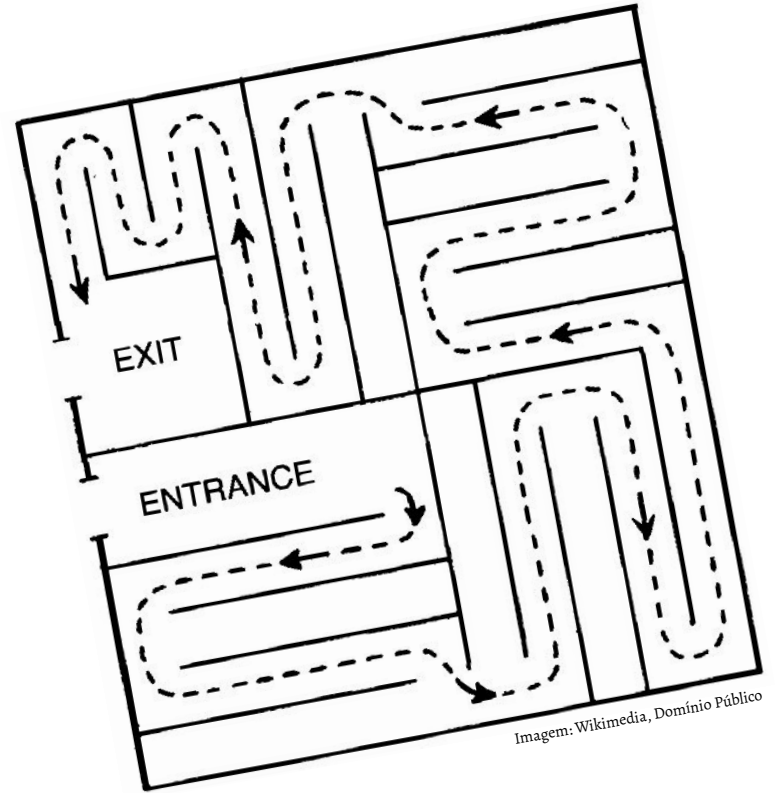
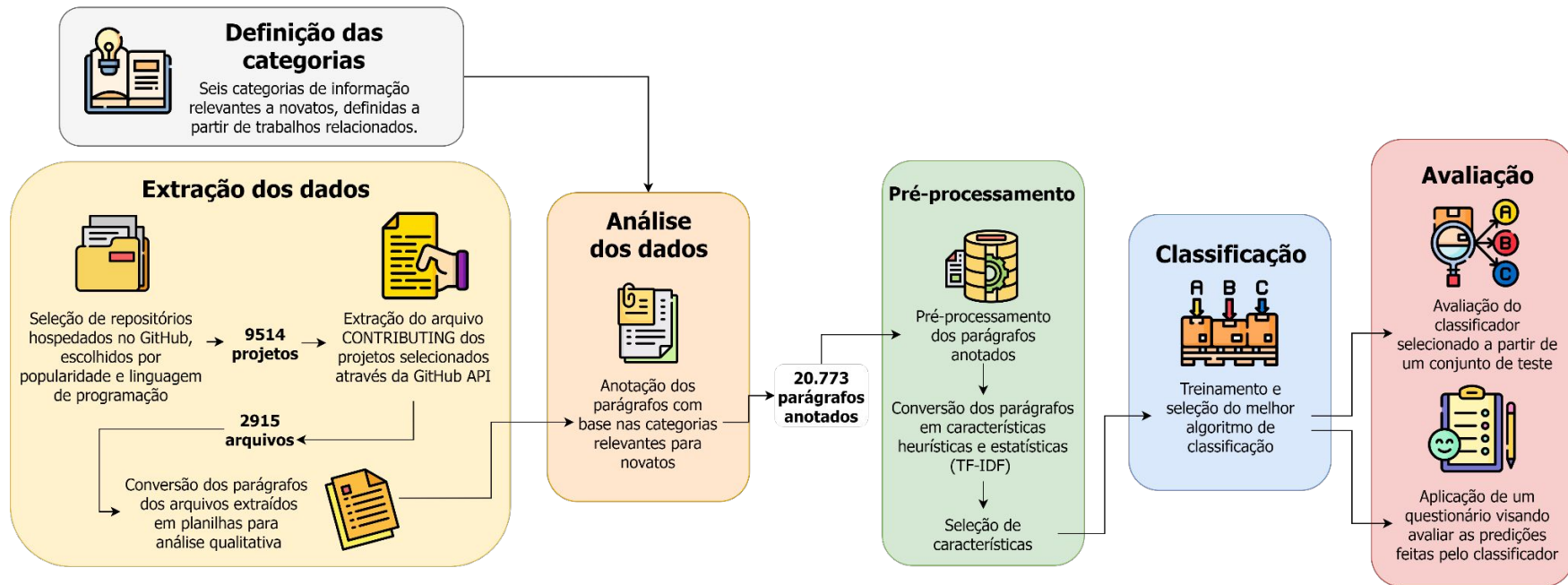


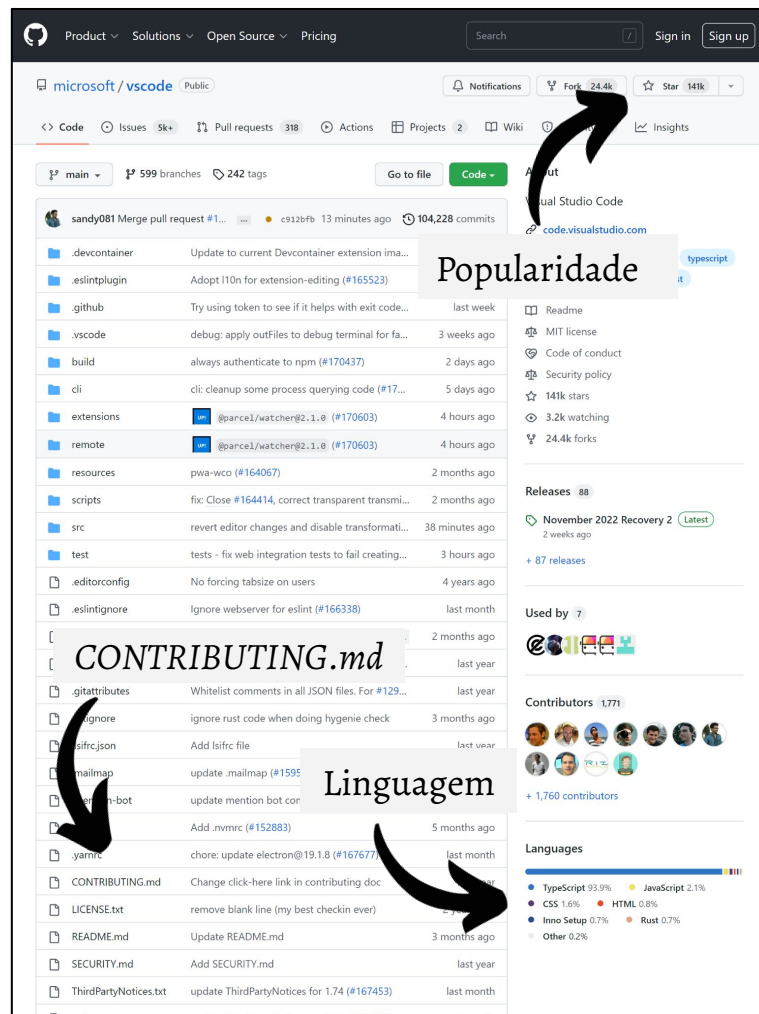
Imagem: Wikimedia, Domínio Público

Método



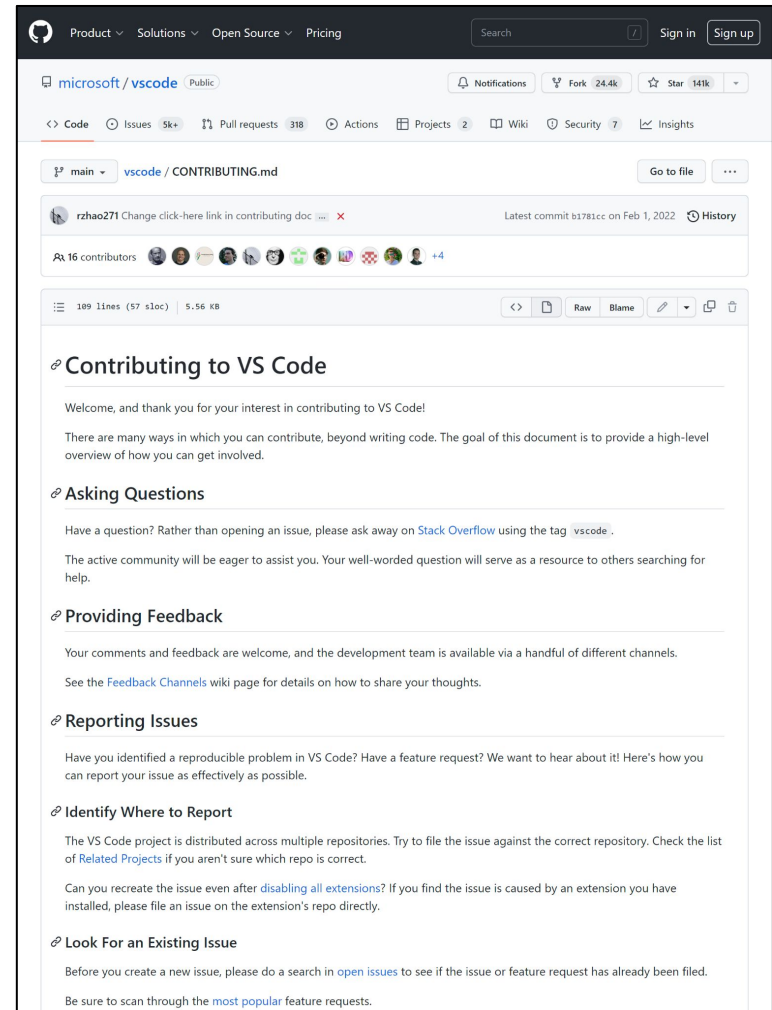
Seleção dos projetos

- 9.514 projetos hospedados no GitHub;
- Filtragem feita por linguagem e popularidade dos projetos;
- Extração do arquivo *CONTRIBUTING.md* via API.



Extração dos arquivos

- *CONTRIBUTING.md* de 2.915 projetos coletados;
- Projetos em que o arquivo não estava em inglês, Markdown ou era < 0.5kb, foram removidos.



Conversão para planilhas

- *Arquivos CONTRIBUTING.md divididos em parágrafos e convertidos em planilhas.*



Product Solutions Open Source Pricing

Search Sign in Sign up

microsoft / vscode Public

Notifications Fork 24.4k Star 141k

<> Code Issues 5k+ Pull requests 318 Actions Projects 2 Wiki Security 7 Insights

main vscode / CONTRIBUTING.md Go to file

rzhao271 Change click-here link in contributing doc Latest commit b1781cc on Feb 1, 2022 History

16 contributors

189 lines (57 sloc) 5.56 KB

Contributing to VS Code

Welcome, and thank you for your interest in contributing to VS Code!

There are many ways in which you can contribute, beyond writing code. The goal of this document is to provide a high-level overview of how you can get involved.

Asking Questions

Have a question? Rather than opening an issue, please ask away on [Stack Overflow](#) using the tag `vscode`.

The active community will be eager to assist you. Your well-worded question will serve as a resource to others searching for help.

Providing Feedback

Your comments and feedback are welcome, and the development team is available via a handful of different channels.

See the [Feedback Channels](#) wiki page for details on how to share your thoughts.

Reporting Issues

Have you identified a reproducible problem in VS Code? Have a feature request? We want to hear about it! Here's how you can report your issue as effectively as possible.

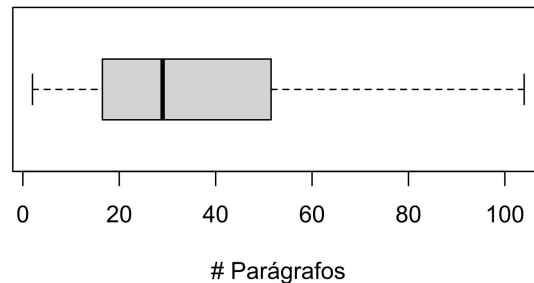
	CF – Contribution flow	CT – Choose a task	TC – Talk to the community	BW – Build local workspace	DC – Deal with the code	SC – Submit the changes
Contributing to VS Code						
Welcome, and thank you for your interest in contributing to VS Code!						
There are many ways in which you can contribute, beyond writing code. The goal of this document is to provide a high-level overview of how you can get involved.						
Asking Questions						

Análise qualitativa

- 30 planilhas foram avaliadas em três etapas por dois pesquisadores. 74.78% de concordância;
- 500 planilhas analisadas separadamente em seguida, resultando em 20.733 parágrafos anotados.

	CF – Contribution flow	CT – Choose a task	TC – Talk to the community
Contributing to Cube.js			
Thanks for taking the time for contribution to Cube.js!			
We're very welcoming community and while it's very Code of Conduct			x
This project and everyone participating in it is governed by the Cube.js Code of Conduct.			
By participating, you are			x
Contributing Code Changes	x		
Please review the preceding section before proposing a code	x		
When you contribute code, you affirm that the contribution is your original work and that you			
Contribution Prerequisites			
Cube.js works with Node.js 8+ and uses yarn as a package			
Development Workflow	x		
Cube.js Client			
1. After cloning Cube.js repository run \$ yarn in packages/cubejs-client-core and			
2. Use \$ yarn link to add these			
3. Perform required code	x		
root to build CommonJS and UMD			

Número de parágrafos analisados por arquivo



Observações

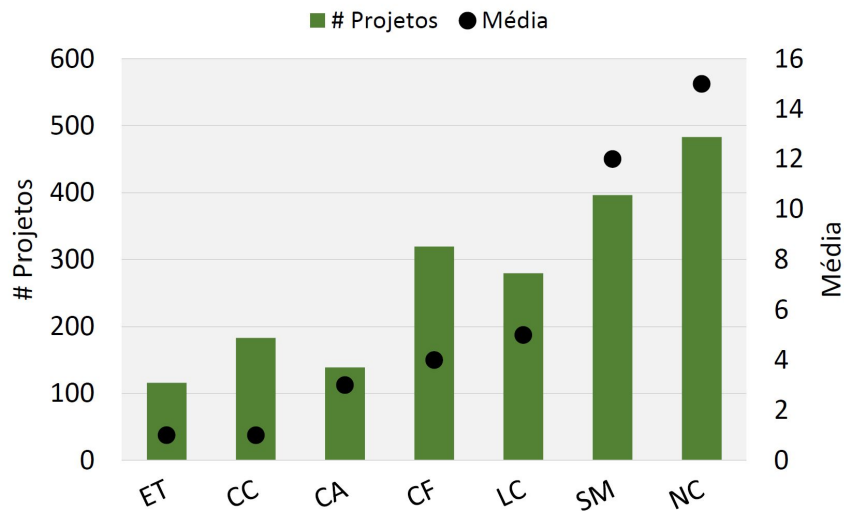


Figura 1. Número de projetos contendo cada categoria de informação, e média de parágrafos em determinada categoria por projeto.

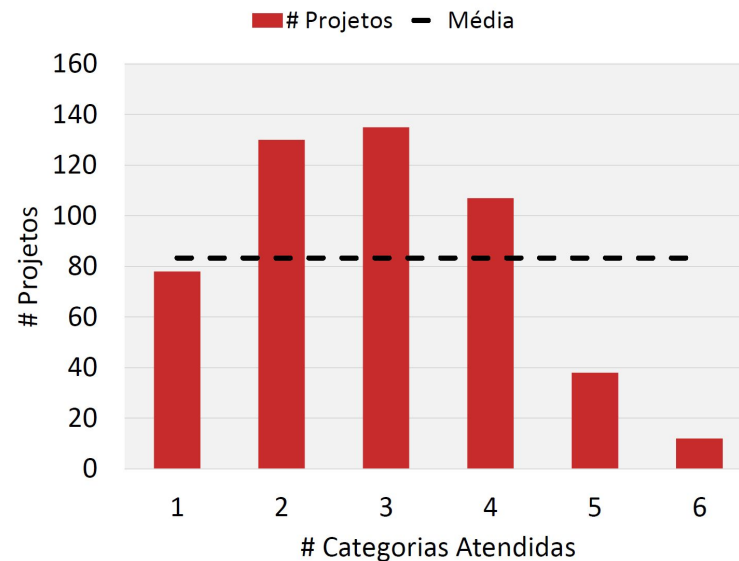
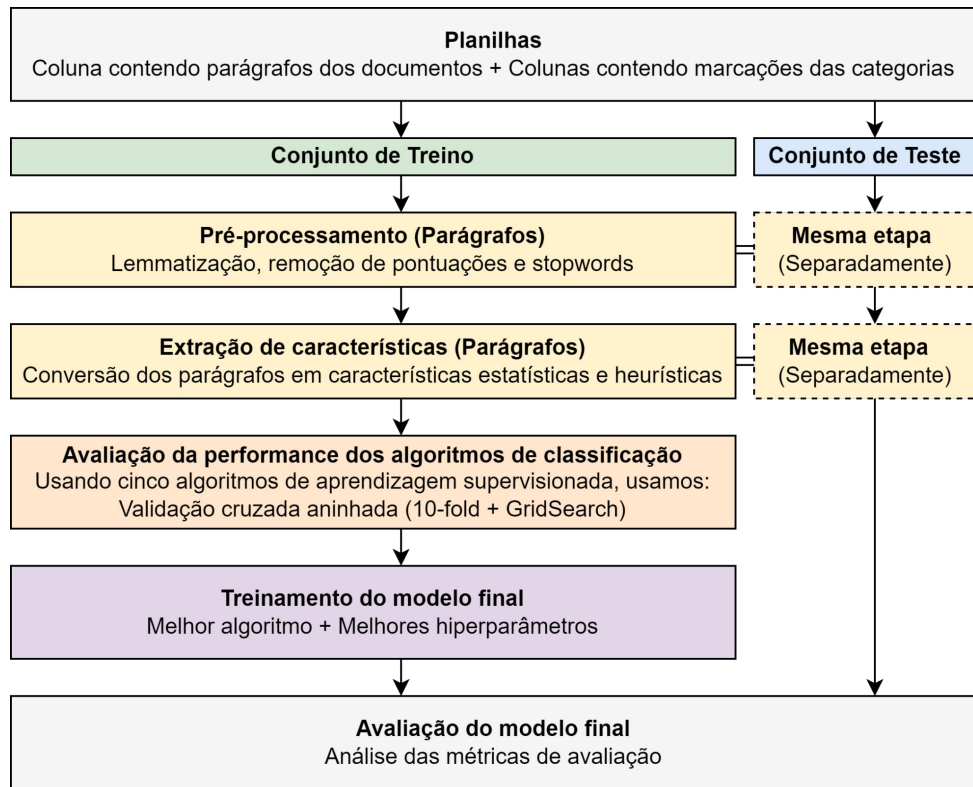


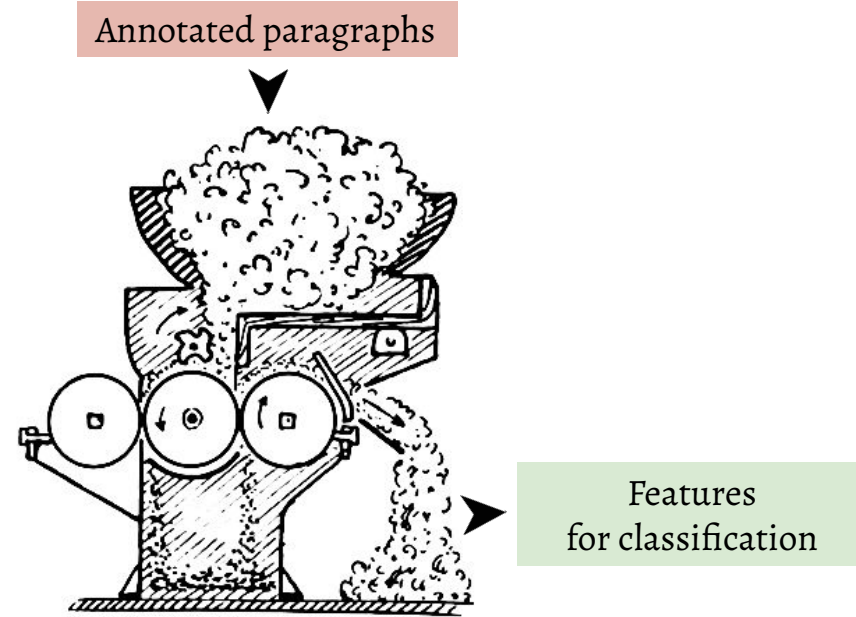
Figura 2. Número de categorias de informação por projeto, com linha média de 83 projetos.

Construção do modelo



Pré-processamento

- Lematização, remoção de pontuações e stopwords via NLTK;
- Conversão dos parágrafos em características heurísticas e estatísticas via *scikit-learn*;
- Seleção de características usando *SelectPercentile*;



Treinamento

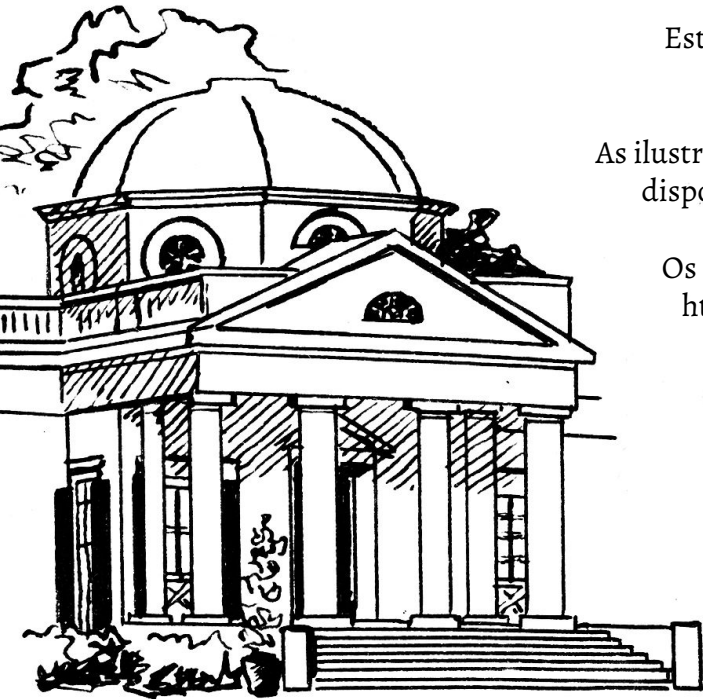


Obrigado !

Este trabalho só foi possível graças ao suporte financeiro da FAPESP (18/02596-1).

As ilustrações utilizadas são cortesia da Scott Foresman, disponíveis em domínio público pela Wikimedia.

Os códigos deste estudo estão disponíveis em:
<https://github.com/fronchetti/dissertacao/>



USP