



UNIVERSITÀ DEGLI STUDI DELL'AQUILA

**Dipartimento di Ingegneria e Scienze dell'Informazione
e Matematica**

Tesi di Laurea Triennale in Informatica

Generating Pull Request Messages with LLMs Using Metadata and Code Changes

Relatore

Prof. Juri Di Rocco

Correlatore

Dr. Claudio Di Sipio

Laureando

Antonio Addario

280548

Anno Accademico 2023-2024

Abstract

This thesis explores the automatic generation of pull request messages in distributed version control systems using a large linguistic model (LLM). The goal is to develop a dataset and a modeling approach capable of predicting and generating effective messages to describe code changes, based on metadata and details of the changes made. The methodology involves extracting pull requests from open-source projects written in Java on GitHub, storing them in a NoSQL database (MongoDB). Subsequently, Ollama was chosen as the LLM, and a Retrieval-Augmented Generation (RAG) approach to structure a scalable and robust prompt that combines the information collected from the dataset in order to generate accurate descriptions of pull requests.

Contents

List of Figures	iv
List of Tables	v
1 Introduction	1
2 Background	2
2.1 Github	2
2.2 Pull request	2
2.3 Large language model	3
2.4 Use case	3
3 Pipeline	5
4 Validation	6
4.1 Configurations	6
4.2 Metrics	6
4.3 Dataset	6
4.4 results	6
5 Related works	7
6 Conclusions	8
bibliography	9
A Appendix	10

List of Figures

2.1	Caption	4
2.2	Caption	4

List of Tables

Chapter 1

Introduction

Chapter 2

Background

This chapter presents the section levels that can be used in the template.

2.1 Github

GitHub is a web-based platform that leverages Git, a distributed version control system, for IT project management and collaboration. The platform is widely used for source control and developer collaboration, making it easier to work in teams on software projects of any scale. GitHub combines Git features, such as version management and distributed control, with an intuitive user interface designed to make it easy for developers of all levels to adopt. In addition, the platform offers a variety of features that improve collaboration and project management, such as access controls to manage who can view or contribute to projects, bug tracking to report and track errors, and feature requests to propose and discuss new ideas. It provides task management tools, centralized documentation through wikis, and robust *pull request* management, which facilitate code reviews and secure integration of changes. Thanks to these characteristics, GitHub is not only a technical tool, but also an ecosystem that fosters collaboration, transparency and innovation, making it one of the most popular platforms among developers and organizations around the world.

2.2 Pull request

A pull request is an essential feature for submitting contributions to a software project, especially in collaborative and open source development contexts. Through an intuitive interface, it allows a contributor to notify changes made to a project's source code, allowing other team members to review, discuss, and integrate the changes in a controlled manner, minimizing the risk of anomalies. However, the pull request is much more than a simple notification, it is a forum dedicated to discussing the proposed functionality. If there are issues with the changes, team members can provide feedback directly within the pull request and even intervene, modifying the functionality via follow-up commits. All of this activity, from reviews to comments to code updates, is tracked and centralized within the pull request, making it a fundamental tool for collaborative and transparent change management. Pull requests also include crucial information, such as associated commits, related issues, and code changes (diffs), ensuring a structured and efficient review.

2.3 Large language model

A large language model (LLM) is an advanced type of language model designed to understand and generate text in complex and general contexts. Its main operation is the processing of huge amounts of text data, which allows the model to learn billions of parameters during the training phase. These models use advanced machine learning techniques, such as transformative neural networks, to generate coherent and relevant responses. However, their development and operation require significant computational resources, both in terms of training time and execution time, making them particularly expensive and complex to manage. Due to their versatility, large language models find application in numerous fields, such as content generation, machine translation, and programming support, demonstrating their usefulness in tackling complex natural language problems.

2.4 Use case

```
octave:1> function xdot = f (x, t)
>
>   r = 0.25; k = 1.4;
>   a = 1.5; b = 0.16; c = 0.9; d = 0.8;
>
>   xdot(1) = r*x(1)*(1 - x(1)/k) - a*x(1)*x(2)/(1 + b*x(1));
>   xdot(2) = c*a*x(1)*x(2)/(1 + b*x(1)) - d*x(2);
>
> endfunction
```

Donec urna leo, vulputate vitae porta eu, vehicula blandit libero. Phasellus eget massa et leo condimentum mollis. Nullam molestie, justo at pellentesque vulputate, sapien velit ornare diam, nec gravida lacus augue non diam. Integer mattis lacus id libero ultrices sit amet mollis neque molestie. Integer ut leo eget mi volutpat congue. Vivamus sodales, turpis id venenatis placerat, tellus purus adipiscing magna, eu aliquam nibh dolor id nibh. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed cursus convallis quam nec vehicula. Sed vulputate neque eget odio fringilla ac sodales urna feugiat.

- Item 1
- Item 2

Donec urna leo, vulputate vitae porta eu, vehicula blandit libero. Phasellus eget massa et leo condimentum mollis. Nullam molestie, justo at pellentesque vulputate, sapien velit ornare diam, nec gravida lacus augue non diam. Integer mattis lacus id libero ultrices sit amet mollis neque molestie. Integer ut leo eget mi volutpat congue. Vivamus sodales, turpis id venenatis placerat, tellus purus adipiscing magna, eu aliquam nibh dolor id nibh. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed cursus convallis quam nec vehicula. Sed vulputate neque eget odio fringilla ac sodales urna feugiat.

1. Item 1
2. Item 2



Figure 2.1: Caption ...



Figure 2.2: Caption ...

Name	Command
Chapter	<code>\chapter{Chapter name}</code>
Section	<code>\section{Section name}</code>
Subsection	<code>\subsection{Subsection name}</code>
Subsubsection	<code>\subsubsection{Subsubsection name}</code>
Paragraph	<code>\paragraph{Paragraph name}</code>
Subparagraph	<code>\paragraph{Subparagraph name}</code>

Chapter 3

Pipeline

Chapter 4

Validation

4.1 Configurations

4.2 Metrics

4.3 Dataset

4.4 results

Chapter 5

Related works

Chapter 6

Conclusions

Paragraph

Subparagraph

Bibliography

Appendix A

Appendix

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus at pulvinar nisi. Phasellus hendrerit, diam placerat interdum iaculis, mauris justo cursus risus, in viverra purus eros at ligula. Ut metus justo, consequat a tristique posuere, laoreet nec nibh. Etiam et scelerisque mauris. Phasellus vel massa magna. Ut non neque id tortor pharetra bibendum vitae sit amet nisi. Duis nec quam quam, sed euismod justo. Pellentesque eu tellus vitae ante tempus malesuada. Nunc accumsan, quam in congue consequat, lectus lectus dapibus erat, id aliquet urna neque at massa. Nulla facilisi. Morbi ullamcorper eleifend posuere. Donec libero leo, faucibus nec bibendum at, mattis et urna. Proin consectetur, nunc ut imperdiet lobortis, magna neque tincidunt lectus, id iaculis nisi justo id nibh. Pellentesque vel sem in erat vulputate faucibus molestie ut lorem.

Quisque tristique urna in lorem laoreet at laoreet quam congue. Donec dolor turpis, blandit non imperdiet aliquet, blandit et felis. In lorem nisi, pretium sit amet vestibulum sed, tempus et sem. Proin non ante turpis. Nulla imperdiet fringilla convallis. Vivamus vel bibendum nisl. Pellentesque justo lectus, molestie vel luctus sed, lobortis in libero. Nulla facilisi. Aliquam erat volutpat. Suspendisse vitae nunc nunc. Sed aliquet est suscipit sapien rhoncus non adipiscing nibh consequat. Aliquam metus urna, faucibus eu vulputate non, luctus eu justo.

Donec urna leo, vulputate vitae porta eu, vehicula blandit libero. Phasellus eget massa et leo condimentum mollis. Nullam molestie, justo at pellentesque vulputate, sapien velit ornare diam, nec gravida lacus augue non diam. Integer mattis lacus id libero ultrices sit amet mollis neque molestie. Integer ut leo eget mi volutpat congue. Vivamus sodales, turpis id venenatis placerat, tellus purus adipiscing magna, eu aliquam nibh dolor id nibh. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed cursus convallis quam nec vehicula. Sed vulputate neque eget odio fringilla ac sodales urna feugiat.

Phasellus nisi quam, volutpat non ullamcorper eget, congue fringilla leo. Cras et erat et nibh placerat commodo id ornare est. Nulla facilisi. Aenean pulvinar scelerisque eros eget interdum. Nunc pulvinar magna ut felis varius in hendrerit dolor accumsan. Nunc pellentesque magna quis magna bibendum non laoreet erat tincidunt. Nulla facilisi.

Duis eget massa sem, gravida interdum ipsum. Nulla nunc nisl, hendrerit sit amet commodo vel, varius id tellus. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ac dolor est. Suspendisse ultrices tincidunt metus eget accumsan. Nullam facilisis, justo vitae convallis sollicitudin, eros augue malesuada metus, nec sagittis diam nibh ut sapien. Duis blandit lectus vitae lorem aliquam nec euismod nisi volutpat. Vestibulum ornare dictum tortor, at faucibus justo tempor non.

Nulla facilisi. Cras non massa nunc, eget euismod purus. Nunc metus ipsum, euismod a consectetur vel, hendrerit nec nunc.

Quisque tristique urna in lorem laoreet at laoreet quam congue. Donec dolor turpis, blandit non imperdiet aliquet, blandit et felis. In lorem nisi, pretium sit amet vestibulum sed, tempus et sem. Proin non ante turpis. Nulla imperdiet fringilla convallis. Vivamus vel bibendum nisl. Pellentesque justo lectus, molestie vel luctus sed, lobortis in libero. Nulla facilisi. Aliquam erat volutpat. Suspendisse vitae nunc nunc. Sed aliquet est suscipit sapien rhoncus non adipiscing nibh consequat. Aliquam metus urna, faucibus eu vulputate non, luctus eu justo.

Donec urna leo, vulputate vitae porta eu, vehicula blandit libero. Phasellus eget massa et leo condimentum mollis. Nullam molestie, justo at pellentesque vulputate, sapien velit ornare diam, nec gravida lacus augue non diam. Integer mattis lacus id libero ultrices sit amet mollis neque molestie. Integer ut leo eget mi volutpat congue. Vivamus sodales, turpis id venenatis placerat, tellus purus adipiscing magna, eu aliquam nibh dolor id nibh. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed cursus convallis quam nec vehicula. Sed vulputate neque eget odio fringilla ac sodales urna feugiat.

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Name Familyname, Rome, Month Year

