

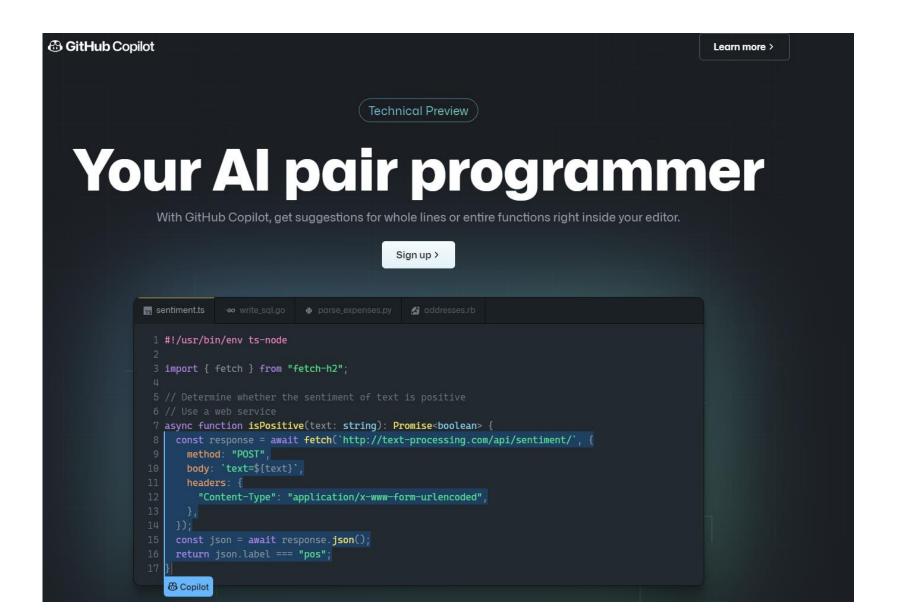
"Mineração de repositórios do Github para construir Sistemas de Recomendação em Engenharia de Software"

Carlos Eduardo Dantas

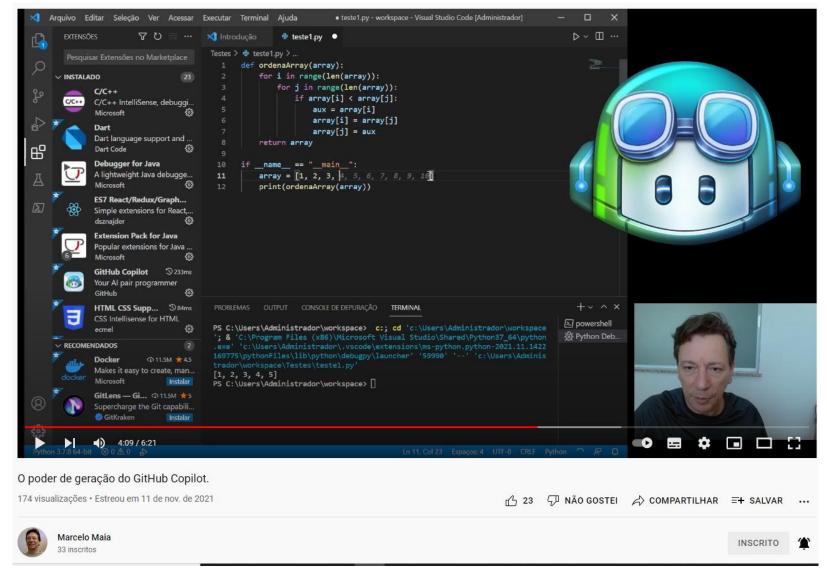
carloseduardodantas@iftm.edu.br



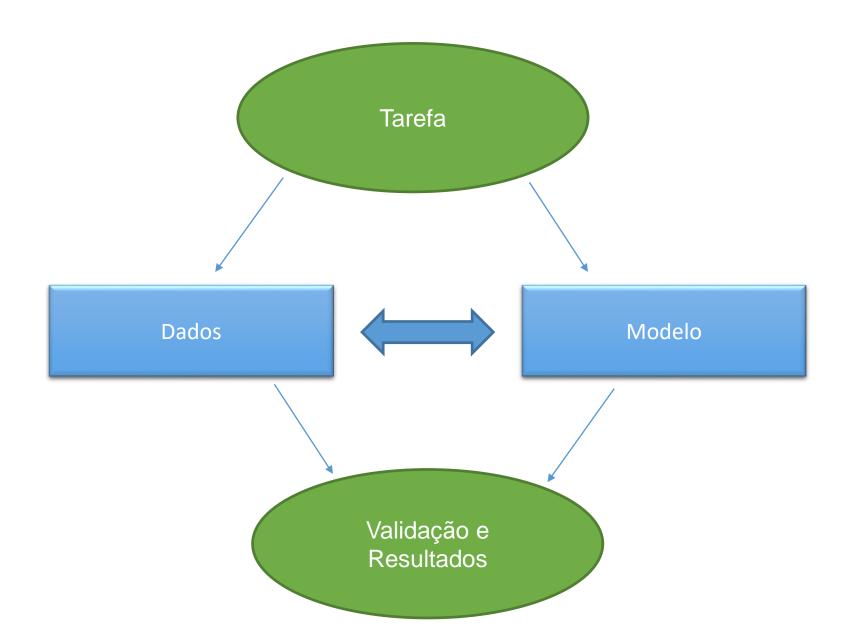
MOTIVAÇÃO



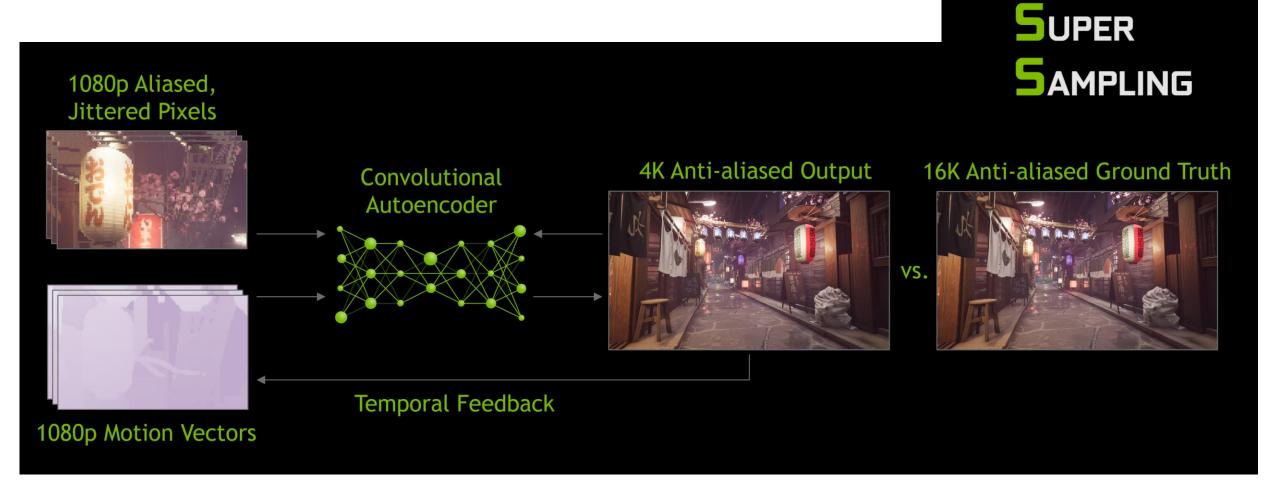
MOTIVAÇÃO



DESIGN

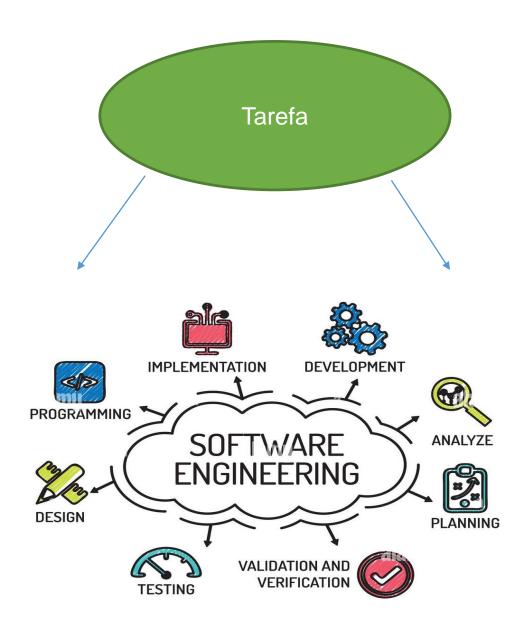


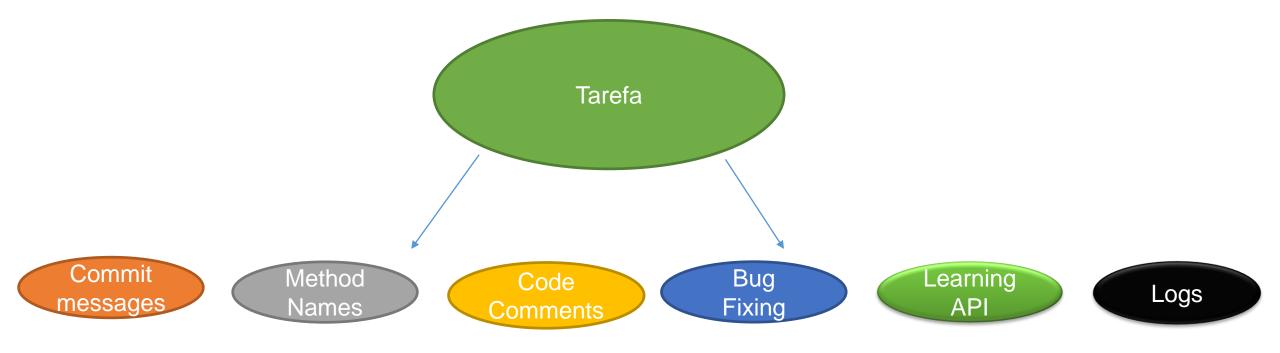
EXEMPLO



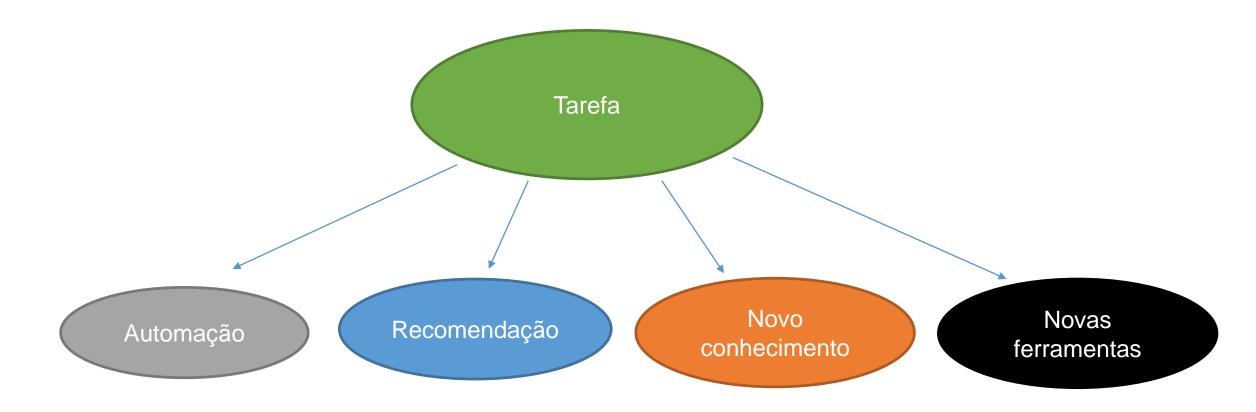
DEEP

LEARNING





AND MANY OTHERS





Detecting Architecturally Relevant Classes Using Dynamic Analysis

Liliane do Nascimento Vale^{1,2} and Marcelo de Almeida Maia¹

¹Faculty of Computing - Federal University of Uberlândia, Uberlândia - MG - Brazil

²Computer Science Department - Federal University of Goiás, Catalão - GO - Brazil

Email: liliane.ufg@gmail.com, marcelo.maia@ufu.br

AutoComment: Mining Question and Answer Sites for Automatic Comment Generation

Edmund Wong, Jinqiu Yang, and Lin Tan University of Waterloo, Waterloo, Ontario, Canada {e32wong, j223yang, lintan}@uwaterloo.ca

Suggesting Accurate Method and Class Names

Miltiadis Allamanis†

Earl T. Barr

Christian Bird*

Charles Sutton

†School of Informatics University of Edinburgh Edinburgh, EH8 9AB, UK {m.allamanis, csutton}@ed.ac.uk

*Dept. of Computer Science University College London London, UK e.barr@ucl.ac.uk *Microsoft Research Microsoft Redmond, WA, USA cbird@microsoft.com

Commit Message Generation for Source Code Changes

Shengbin Xu¹, Yuan Yao¹, Feng Xu¹, Tianxiao Gu², Hanghang Tong³ and Jian Lu¹

¹State Key Laboratory for Novel Software Technology, Nanjing University, China

²Alibaba Group, USA

³Arizona State University, USA

kingxu@smail.nju.edu.cn, {y.yao, xf, lj}@nju.edu.cn, tianxiao.gu@gmail.com, hanghang.tong@asu.edu



ICSE-2019-AUTOFIX Mer	ge pull request #4 from martinezmatias/patchesinfo	37e0447 on 30 Oct 2020 🖰 5 commits
csc csc	Fixes	2 years ago
results	info about the patches	13 months ago
tl tl	Fixes	2 years ago
README.md	upload readme	17 months ago
main.py	Fixes	2 years ago
∷≣ README.md		

DLFix: Context-based Code Transformation Learning for Automated Program Repair



When and Why Your Code Starts to Smell Bad

Michele Tufano*, Fabio Palomba[†], Gabriele Bavota[‡], Rocco Oliveto[§],

Massimiliano Di Penta[¶], Andrea De Lucia[†], Denys Poshyvanyk*

*The College of William and Mary, Williamsburg, VA, USA - [†]University of Salerno, Fisciano (SA), Italy

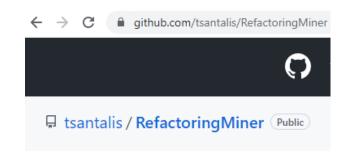
[‡]Free University of Bozen-Bolzano, Italy - [§]University of Molise, Pesche (IS), Italy

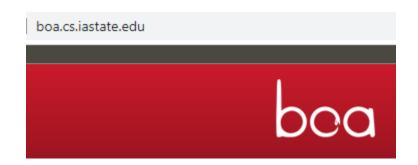
¶University of Sannio, Benevento, Italy

Why We Refactor? Confessions of GitHub Contributors

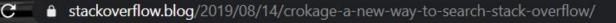
Danilo Silva Universidade Federal de Minas Gerais, Brazil danilofs@dcc.ufmg.br Nikolaos Tsantalis Concordia University Montreal, Canada tsantalis@cse.concordia.ca Marco Tulio Valente Universidade Federal de Minas Gerais, Brazil mtov@dcc.ufmg.br











Company



Podcast

Essays, opinions, and advice on the act of computer programming from Stack Overflow.



Benjamin Popper · 3° Director of Content at Stack Overflow Brooklyn, Nova York, Estados Unidos · 346 conexões · Informações de contato code-for-a-living

Latest

Newsletter

AUGUST 14, 2019

CROKAGE: A New Way to Search Stack Overflow

One of the most powerful attributes of Stack Overflow (SO) is the accumulation of developers' knowledge over time. Community members have contributed more than 18 million questions and 27 million answers. When a developer is stuck on a coding problem, they search through this vast trove of information to see if a solution to their...



apisonar.com

APISonar

APISonar: Mining API usage examples

Andre Hora®

Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte, Brazil

Correspondence

Andre Hora, Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte, Brazil. Email: andrehora@dcc.ufmg.br

Abstract

Developers spend a significant part of their time searching for code examples on the web. Often, they look for Application Programming Interface (API) usage examples, that is, how to use APIs provided by libraries and frameworks. For this purpose, several programming websites are available. Some programming websites provide manually created examples: unfortunately, as millions of APIs are available nowadays, they do not cover the majority of the APIs. To alleviate this limitation, other programming websites focus on automatically mining API usage examples from code repositories. To the best of our knowledge, however, these solutions are still very limited: they often present poor, duplicated, and similar API usage examples. In this article, we propose an approach, APISonar, to automatically mine API usage examples from code repositories. Our approach aims to overcome the limitations of current solutions: we focus on presenting readable and reusable API usage examples. We analyze millions of source files provided by 4486 software projects hosted on GitHub. Based on this data, we extract 11 million API usage examples about 1.5 million distinct APIs. We evaluate APISonar by assessing its quality and usage. We show that APISonar is a competitive solution, providing the best API examples in terms of readability and reusability, as compared with popular programming websites. Moreover, despite being a novel website, APISonar attracted a significant amount of users in a short period (3.7K users from 119 countries during 5 months). APISonar is available at www.apisonar.com.

Search API examples

try: java, android, Log, math, json...

Learning to Spot and Refactor Inconsistent Method Names

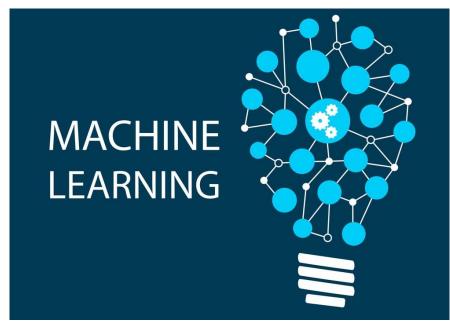
Kui Liu[†], Dongsun Kim[†], Tegawendé F. Bissyandé[†], Taeyoung Kim[‡], Kisub Kim[†], Anil Koyuncu[†], Suntae Kim[‡], Yves Le Traon[†]

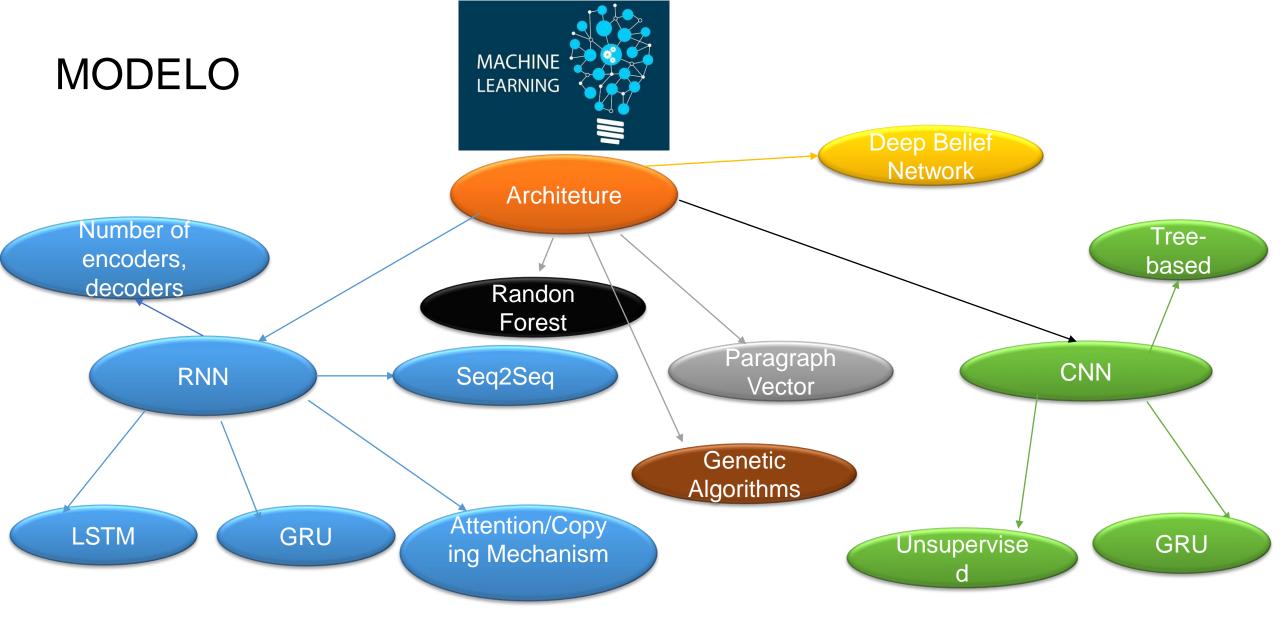
†Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg {kui.liu, dongsun.kim, tegawende.bissyande, kisub.kim, koyuncu.anil, yves.letraon}@uni.lu †Department of Software Engineering, Chonbuk National University, South Korea {rlaxodud1200, jipsin08}@gmail.com

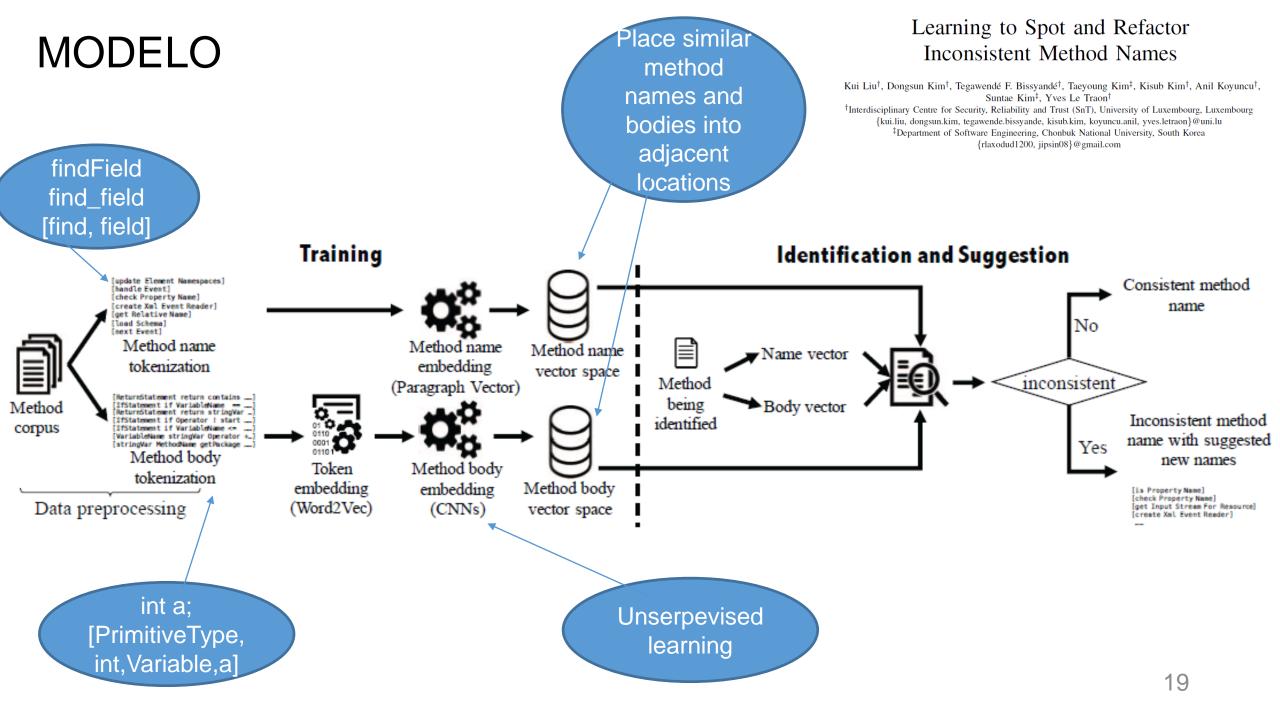
Commit 70106770ea61a5fe845653a0b793f4934cc00144
-public double inverseCummulativeProbability(final double p){
+public double inverseCumulativeProbability(final double p){

MODELO









MODELO

← → C

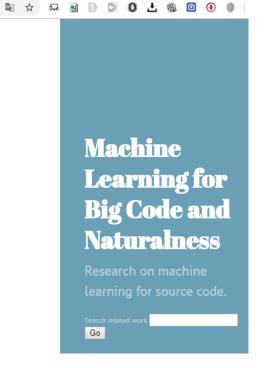
ml4code.github.io/papers.html

work.

Papers from https://ml4code.github.io/papers.html



	Search:				
Year ▼	Title	Authors			
2019	A case study on machine learning for synthesizing benchmarks	A. Goens, A. Brauckmann, S. Ertel, C. Cummins, H. Leather, J. Castrillon			
2019	A Grammar-Based Structural CNN Decoder for Code Generation	Z. Sun, Q. Zhu, L. Mou, Y. Xiong, G. Li, L. Zhang			
2019	A Literature Study of Embeddings on Source Code	Z. Chen, M. Monperrus			
2019	A Neural Model for Generating Natural Language Summaries of Program Subroutines	A. LeClair, S. Jiang, C. McMillan			
2019	A Neural Model for Method Name Generation from Functional Description	S. Gao, C. Chen, Z. Xing, Y. Ma, W. Song, S.W. Lin			

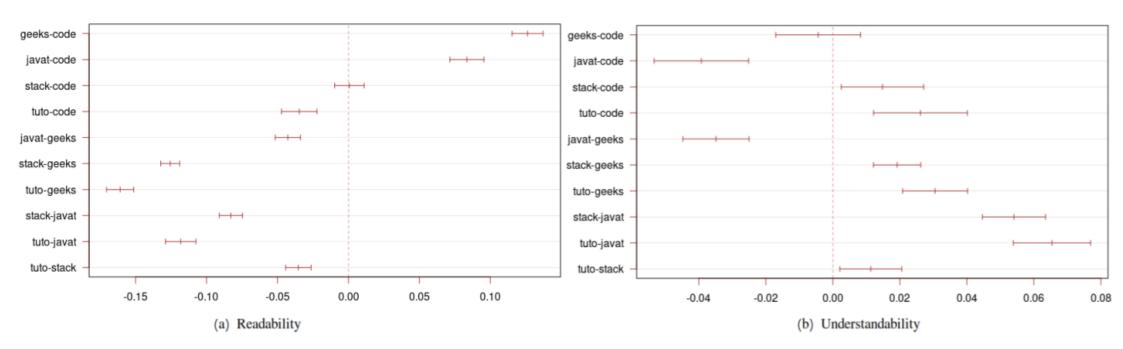


MODELO

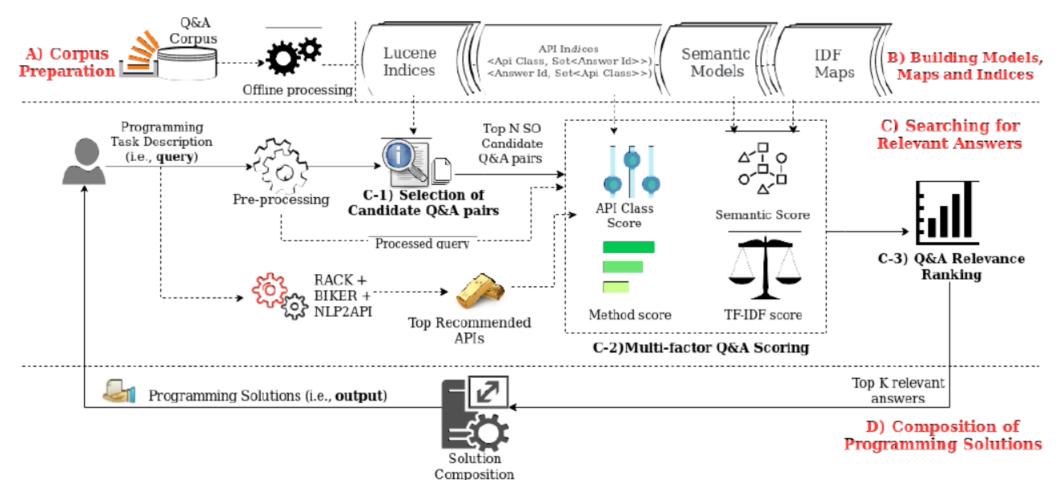
Readability and Understandability of Snippets Recommended by General-purpose Web Search Engines: a Comparative Study



Marcelo A. Maia marcelo.maia@ufu.br Federal University of Uberlândia Brazil







https://archive.org/details/stackexchange



Stack Exchange Data Dump

by Stack Exchange, Inc.

Publication date 2021-09-07

Usage Attribution-ShareAlike 4.0 International @ (1)

Topics Stack Exchange Data Dump
Contributor Stack Exchange Community

This is an anonymized dump of all user-contributed content on the Stack Exchange network. Each site is formatted as a separate archive consisting of XML files zipped via 7-zip using bzip2 compression. Each site archive includes Posts, Users, Votes, Comments, PostHistory and PostLinks. For complete schema information, see the included readme.txt.

All user content contributed to the Stack Exchange network is cc-by-sa 4.0 licensed, intended to be shared and remixed. We even provide all our data as a convenient data dump.

License: https://creativecommons.org/licenses/by-sa/4.0/

stackoverflow.com-Badges.7z	296.4M
stackoverflow.com-Comments.7z	4.8G
stackoverflow.com-PostHistory.7z	29.7G
stackoverflow.com-PostLinks.7z	103.3M
stackoverflow.com-Posts.7z	16.9G
stackoverflow.com-Tags.7z	867.3K
stackoverflow.com-Users.7z	777.9M
stackoverflow.com-Votes.7z	1.3G

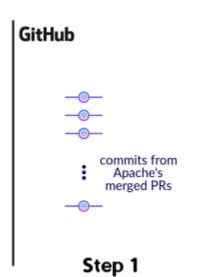
An Empirical Study on Refactoring-Inducing Pull Requests

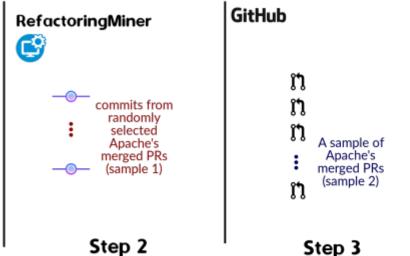
Flávia Coelho Federal University of Campina Grande Campina Grande, Brazil flavia@copin.ufcg.edu.br

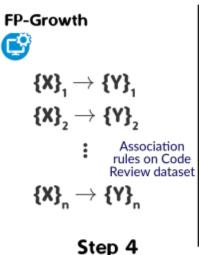
Tiago Massoni Federal University of Campina Grande Campina Grande, Brazil massoni@computacao.ufcg.edu.br Nikolaos Tsantalis Concordia University Montreal, Canada nikolaos.tsantalis@concordia.ca

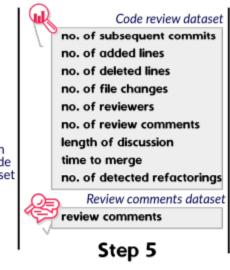
Everton L. G. Alves
Federal University of Campina Grande
Campina Grande, Brazil
everton@computacao.ufcg.edu.br











Mining of raw pull requests data

Mining the merged PRs from Apache's repositories

Output

Pull requests dataset Commits dataset

Detection of refactoring edits

Listing of refactoring edits mined in the commits from a randomly selected Apache's merged PRs

Output

Refactorings dataset

Mining of raw code review data

Mining of code review data from a sample of Apache's PRs obtained based on the Refactorings dataset

Output

Code review dataset Review comments dataset

Association rule learning

ARL on the quantitative features from the code review dataset

> Output Hypotheses

Data analysis

Statistical testing of hypotheses, and manual inspection of review comments

Why We Refactor? Confessions of GitHub Contributor



Danilo Silva Universidade Federal de Minas Gerais, Brazil danilofs@dcc.ufmg.br Nikolaos Tsantalis Concordia University Montreal, Canada tsantalis@cse.concordia.ca Marco Tulio Valente Universidade Federal de Minas Gerais, Brazil mtov@dcc.ufmg.br

```
public static List<CourseInfo> getCourses() {
                                                                                     public static List<CourseInfo> getCourses() {
                                                                                       try {
   List<CourseInfo> result = new ArrayList<CourseInfo>();
                                                                                         List<CourseInfo> result = new ArrayList<CourseInfo>();
   final List<CourseInfo> courseInfos =
                                                                                         int pageNumber = 0;
                                                                                         boolean hasNext = addCoursesFromStepic(result, pageNumber)
         getFromStepic("courses", CoursesContainer.class).courses;
                                                                                         while (hasNext) {
   for (CourseInfo info : courseInfos) {
     final String courseType = info.getType();
                                                                                           pageNumber += 1;
     if (StringUtil.isEmptyOrSpaces(courseType)) continue;
                                                                                           hasNext = addCoursesFromStepic(result, pageNumber)
      final List<String> typeLanguage = StringUtil.split(courseType, " ");
     if (typeLanguage.size() == 2 && PYCHARM_PREFIX.equals(typeLanguage.get(0))) {
                                                                                         return result;
       result.add(info);
                                                                                       catch (IOException e) {
                                                                                         LOG.error("Cannot load course list " + e.getMessage());
   return result;
                                                                                       return Collections.emptyList();
  catch (IOException e) {
   LOG.error("Cannot load course list " + e.getMessage());
                                                                                     private static boolean addCoursesFromStepic(List<CourseInfo> result, int pageNumber)
                                                                                         throws IOException {
  return Collections.emptyList();
                                                                                        final String url = pageNumber == 0 ? "courses" : "courses?page=" +
                                                                                         String.valueOf(pageNumber);
                                                                                        final CoursesContainer coursesContainer = getFromStepic(url, CoursesContainer.class);
                                                                                        final List<CourseInfo> courseInfos = coursesContainer.courses;
                                                                                        for (CourseInfo info : courseInfos) {
                                                                                         final String courseType = info.getType();
           Extracted Code
                                                                                         if (StringUtil.isEmptyOrSpaces(courseType)) continue;
                                                                                         final List<String> typeLanguage = StringUtil.split(courseType, " ");
            Added Code
                                                                                         if (typeLanguage.size() == 2 && PYCHARM PREFIX.equals(typeLanguage.get(0))) {
                                                                                          result.add(info);
           Call to the Extracted Method
                                                                                        return coursesContainer.meta.containsKey("has next") &&
                                                                                         coursesContainer.meta.get("has_next") == Boolean.TRUE
```

An Exploratory Study of Log Placement Recommendation in an Enterprise System

Jeanderson Cândido*†, Jan Haesen†, Maurício Aniche*, and Arie van Deursen*

*Department of Software Technology

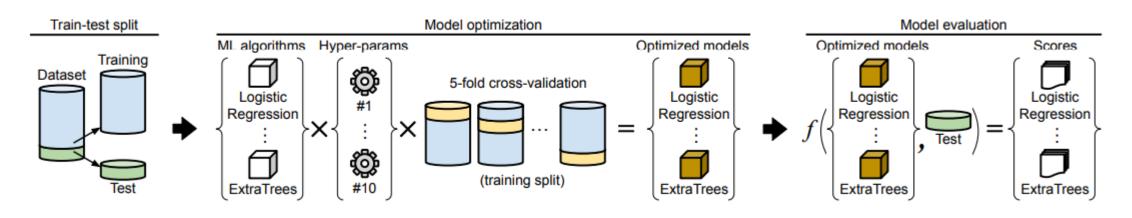
Delft University of Technology, The Netherlands

{j.candido, m.f.aniche, arie.vandeursen}@tudelft.nl

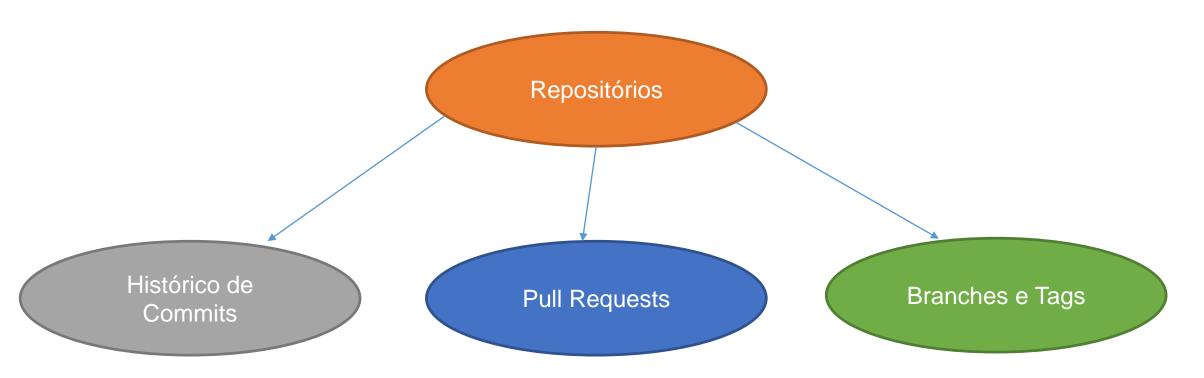
†Adyen N.V., The Netherlands

{jeanderson.candido, jan.haesen}@adyen.com









Quais repositórios selecionar?



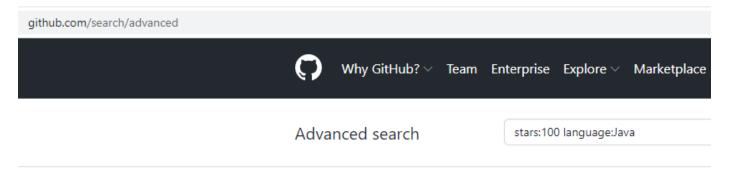
Ecossistemas

Estrelas

Linguagens de Programação Quantidade de commits

Tempo





VALIDAÇÃO

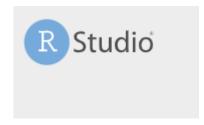
K = 10

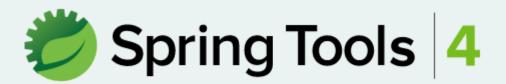
	Java			
	Hit	MRR	MAP	MR
BIKER	0.16	0.11	0.11	0.01
BM25 + API Class	0.58	0.18	0.17	0.10
BM25 + Sent2Vec	0.49	0.22	0.20	0.08
BM25	0.56	0.22	0.22	0.13
BM25 + Method	0.72	0.40	0.36	0.16
BM25 + fastText	0.67	0.39	0.34	0.13
BM25 + TF-IDF	0.63	0.34	0.32	0.16
CROKAGE	0.81	0.55	0.49	0.22

K = 1

	Java			
•	Hit	MRR	MAP	MR
BIKER	0.07	0.07	0.07	0.01
BM25 + API Class	0.00	0.00	0.00	0.00
BM25 + Sent2Vec	0.12	0.12	0.12	0.01
BM25	0.09	0.09	0.09	0.01
BM25 + Method	0.25	0.25	0.25	0.02
BM25 + fastText	0.28	0.28	0.28	0.03
BM25 + TF-IDF	0.21	0.21	0.21	0.03
CROKAGE	0.46	0.46	0.46	0.06







Spring Tools 4 is the next generation of Spring tooling for your favorite coding environment. Largely rebuilt from scratch, it provides world-class support for developing Spring-based enterprise applications, whether you prefer Eclipse, Visual Studio Code, or Theia IDE.

https://www.rstudio.com/

https://spring.io/tools

Tarefa

Quais operações de refactoring tendem a melhorar a legibilidade do código-fonte?

Quais operações de refactoring tendem a melhorar a legibilidade do código-fonte?

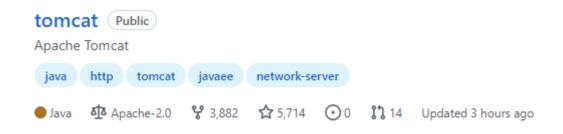
Repositórios do Github

Ferramenta que detecta operações de refactoring

Ferramenta que captura a legibilidade das classes

Análise estatística

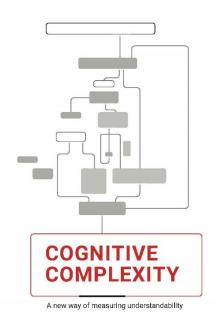
Quais operações de refactoring tendem a melhorar a legibilidade do código-fonte?





Quais operações de refactoring tendem a melhorar a legibilidade do código-fonte?





A comprehensive model for code readability

Simone Scalabrino 🔀 Mario Linares-Vásquez, Rocco Oliveto, Denys Poshyvanyk

First published: 08 June 2018 | https://doi.org/10.1002/smr.1958 | Citations: 10

By G. Ann Campbell, Product Owner - SonarSource SA

Quais operações de refactoring tendem a melhorar a legibilidade do código-fonte?

```
Refactoring-readability - RefactoringReadabilityApplication [Spring Boot App] C:\Program Files\Java\jdk1.8.0_151\bin\javaw.exe (23 de nov de 2021 19:19:31)
2021-11-23 19:26:56.102 INFO 7260 --- [va AST analyzer] o.s.analyzer.commons.ProgressReport
                                                                                                  : 1 source file to be analyzed
2021-11-23 19:26:56.143 INFO 7260 --- [va AST analyzer] o.s.analyzer.commons.ProgressReport
                                                                                                  : 1/1 source file has been analyzed
2021-11-23 19:26:56.143 WARN 7260 ---
                                                   main] org.sonar.java.SonarComponents
                                                                                                  : Unresolved imports/types have been detected during analysis. Enable DEBUG mode to see them.
2021-11-23 19:26:56.144 INFO 7260 ---
                                                   mainl o.r.rm1.GitHistoryRefactoringMinerImpl
                                                                                                  : Processing tomcat [Commits: 10, Errors: 0, Refactorings: 42]
2021-11-23 19:26:56.165 INFO 7260 ---
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
                                                                                                  : Processing F:\repositories\tomcat 5a55179a049ec39246c8ba76073ab4485542b98c ...
                                                                                                  : Processing F:\repositories\tomcat 808e0cb47bc428404e58775606bbe933193340f4 ...
2021-11-23 19:26:56.181 INFO 7260 ---
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
Refactorings at 808e0cb47bc428404e58775606bbe933193340f4
Change Variable Type e: GSSException to e: GSSException | IllegalStateException in method public getUserPrincipal(): Principal from class org.apache.catalina.connector.Request
2021-11-23 19:27:12.299 INFO 7260 --- [va AST analyzer] o.s.analyzer.commons.ProgressReport
                                                                                                  : 1 source file to be analyzed
2021-11-23 19:27:12.414 INFO 7260 --- [va AST analyzer] o.s.analyzer.commons.ProgressReport
                                                                                                  : 1/1 source file has been analyzed
                                                   main] org.sonar.java.SonarComponents
                                                                                                  : Unresolved imports/types have been detected during analysis. Enable DEBUG mode to see them.
2021-11-23 19:27:12.415 WARN 7260 ---
                                                                                                  : 1 source file to be analyzed
2021-11-23 19:27:30.195 INFO 7260 --- [va AST analyzer] o.s.analyzer.commons.ProgressReport
2021-11-23 19:27:30.296 INFO 7260 --- [va AST analyzer] o.s.analyzer.commons.ProgressReport
                                                                                                  : 1/1 source file has been analyzed
                                                                                                  : Unresolved imports/types have been detected during analysis. Enable DEBUG mode to see them.
2021-11-23 19:27:30.296 WARN 7260 ---
                                                   main] org.sonar.java.SonarComponents
2021-11-23 19:27:30.297 INFO 7260 ---
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
                                                                                                  : Processing tomcat [Commits: 11, Errors: 0, Refactorings: 43]
2021-11-23 19:27:30.309 INFO 7260 ---
                                                   mainl o.r.rm1.GitHistorvRefactoringMinerImpl
                                                                                                  : Processing F:\repositories\tomcat f9e8d2b09f9e9a6c02c674c1464fb5abf68fb000 ...
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
                                                                                                  : Processing F:\repositories\tomcat 5a55179a049ec39246c8ba76073ab4485542b98c ...
2021-11-23 19:27:30.317 INFO 7260 ---
Refactorings at 5a55179a049ec39246c8ba76073ab4485542b98c
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
2021-11-23 19:27:30.401 INFO 7260 --- [
                                                                                                  : Processing F:\repositories\tomcat a791804b0951da1a67c026926345d6b7191273fd ...
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
2021-11-23 19:27:30.412 INFO 7260 ---
                                                                                                  : Processing F:\repositories\tomcat f9e8d2b09f9e9a6c02c674c1464fb5abf68fb000 ...
Refactorings at f9e8d2b09f9e9a6c02c674c1464fb5abf68fb000
2021-11-23 19:27:30.502 INFO 7260 ---
                                                   mainl o.r.rm1.GitHistorvRefactoringMinerImpl
                                                                                                  : Processing F:\repositories\tomcat d83479efe35e15125a3be781c9670269abe74bfe ...
2021-11-23 19:27:30.509 INFO 7260 ---
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
                                                                                                 : Processing F:\repositories\tomcat a791804b0951da1a67c026926345d6b7191273fd ...
Refactorings at a791804b0951da1a67c026926345d6b7191273fd
Refactorings at d83479efe35e15125a3be781c9670269abe74bfe
2021-11-23 19:27:30.607 INFO 7260 ---
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl
                                                                                                 : Processing F:\repositories\tomcat c3412ad28f72cbaf024967978c1c1690d3c5a27d ...
2021-11-23 19:27:30.614 INFO 7260 --- [
                                                   main] o.r.rm1.GitHistoryRefactoringMinerImpl : Processing F:\repositories\tomcat 248e9b74c9ccb8fb7b5072acd932f9deb3e979fb ...
Refactorings at 248e9b74c9ccb8fb7b5072acd932f9deb3e979fb
                org.apache.tomcat.util.net.openssl.panama.OpenSSLContext.DHParam moved to org.apache.tomcat.util.net.openssl.panama.OpenSSLLifecycleListener.DHParam
```

Quais operações de refactoring tendem a melhorar a legibilidade do código-fonte?

Α Α	В	C	D	E	F
repository	refactoring_type	className	operation	readability	understandability
https://github.com/apache/tomcat.git	Add Parameter	org.apache.tomcat.util.net.openssl.panama.OpenSSLContext.ContextState	BEFORE	0.5000286049318571	345.0
https://github.com/apache/tomcat.git	Add Parameter	org.apache.tomcat.util.net.openssl.panama.OpenSSLContext.ContextState	AFTER	0.4972928118180621	345.0
https://github.com/apache/tomcat.git	Move Attribute	org.apache.tomcat.util.net.openssl.panama.OpenSSLContext	BEFORE	0.5000286049318571	345.0
https://github.com/apache/tomcat.git	Move Attribute	org.apache.tomcat.util.net.openssl.panama.OpenSSLContext.ContextState	AFTER	0.4972928118180621	345.0
https://github.com/apache/tomcat.git	Change Attribute Access Modifier	org.apache.tomcat.util.net.openssl.panama.OpenSSLContext	BEFORE	0.5000286049318571	345.0
https://github.com/apache/tomcat.git	Change Attribute Access Modifier	org.apache.tomcat.util.net.openssl.panama.OpenSSLContext.ContextState	AFTER	0.4972928118180621	345.0

Quais operações de refactoring tendem a melhorar a legibilidade do código-fonte?

