

Looking for related discussion posts on *GitHub Discussions*

Márcia Lima¹, Igor Steinmacher², Denae Ford³, Evangeline Liu⁴, Grace Vorreuter⁵, Tayana Conte⁶, and Bruno Gadelha⁷

¹Federal University of Amazonas (UFAM) & Amazonas State University (UEA)

²Northern Arizona University (NAU)

³Microsoft Research

⁴GitHub

⁵GitHub

⁶Federal University of Amazonas (UFAM)

⁷Federal University of Amazonas (UFAM)

Corresponding author:

Márcia Lima¹

Email address: msllima@uea.edu.br;marcia.lima@icomp.ufam.edu.br

ABSTRACT

Software teams increasingly adopt different tools and communication channels to aid the software collaborative development model and coordinate tasks. Among such resources, software development forums have become widely used by developers. Such environments enable developers to get and share technical information. Interested in supporting the development and management of Open Source Software (OSS) projects, GitHub announced *GitHub Discussions* — a native forum to facilitate collaborative discussions between users and members of communities hosted on the platform. Since GitHub Discussions is a software development discussion forum, it faces challenges similar to those faced by systems used for asynchronous communication, including related discussion posts (duplicates or near-duplicated posts). While duplicate posts have the same content—and may be exact copies—near-duplicates share similar topics and information. Both can introduce noise to the platform and compromise project knowledge sharing. This paper addresses the problem of detecting related discussion posts on the GitHub Discussions forum. To do so, we propose an approach based on a Sentence-BERT pre-trained model: the *RD-Detector*. We evaluated *RD-Detector* using data from three OSS communities. Our dataset comprises 16,048 discussion posts. Three OSS maintainers and three Software Engineering (SE) researchers manually evaluated the *RD-Detector* results, achieving 77-100% of precision and 66% of recall. In addition, maintainers pointed out practical applications of the approach, such as providing knowledge to support merging the discussion posts and converting the discussion to comments on other related posts. OSS maintainers can benefit from *RD-Detector* to address the labor-intensive task of manually detecting related posts.

DISCUSSIONS

In this section, we discuss the effects of changing the K value and the false-positive predictions.

The impacts of changing the K value

As discussed in the selection of related discussion candidates section, the K value delimits the search bounds for related discussion candidates. As we increase the K value, the search space boundaries also increase. Conversely, as we decrease the K value, the search space boundaries decrease, increasing the chances of detecting duplicates. The K value affects the value of the threshold $T_{related}$ and, consequently, the number of detected related discussion candidates. By configuring the *RD-Detector* to run over $K = 5$ and $K = 10$, we note that the sets of related discussion candidates created when we set $K = 5$ are subsets of $K = 10$. Therefore, there is a risk of false-positive predictions propagation through the related discussion candidates' sets. We identified this propagation problem by analyzing the sets of related discussion candidates detected considering the configuration groups $p = Gatsby$ and $c = Ideas$. The

same unrelated discussion pair (*master, target*) occurs in $R_{p=\text{Gatsby}|c=\text{Ideas}|K=5}$ and $R_{p=\text{Gatsby}|c=\text{Ideas}|K=10}$. Although the precision rate tends to decrease, the approach detects new pairs of related discussions when we vary the K values from 5 to 10.

Gatsby project: Analyzing the sets of candidates $R_{p=\text{Gatsby}|c=\text{Q\&A}|K=5}$ and $R_{p=\text{Gatsby}|c=\text{Q\&A}|K=10}$, RD-Detector detected two new pairs of related discussions when changing the K value from 5 to 10. The SE researchers judged discussions from both new pairs as related. Since all related discussion candidates from both sets were indeed related, the RD-Detector achieved the best precision rate (100%). Considering $p = \text{Gatsby}|c = \text{Ideas}|K = 5$ and $p = \text{Gatsby}|c = \text{Ideas}|K = 10$, the RD-Detector detected three new pairs of related discussion candidates by increasing the K value. However, one out of the three new pairs was judged unrelated by M_Gatsby. In this case, the precision rate decreased from 83.33% to 77.78%.

As discussed in the preprocessing phase applied to discussions dataset section, when we set the category filter to ALL ($c = \text{ALL}$), the RD-Detector calculates the similarity values between all discussion pairs in the dataset. The approach detected seven related discussion candidates using the configuration $p = \text{Gatsby}|c = \text{ALL}|K = 5$ ($R_{p=\text{Gatsby}|c=\text{ALL}|K=5} = 7$), of which five elements are also present in set $R_{p=\text{Gatsby}|c=\text{Q\&A}|K=10}$. Evaluators judged all five elements in $R_{p=\text{Gatsby}|c=\text{Q\&A}|K=10}$ set as related. M_Gatsby judged that the discussions of the two new pairs were related. We analyzed the discussions' content and found that users create related discussions in different categories. The discussions of these two specific pairs were from Idea and Q&A categories, respectively.

Regarding the set $R_{p=\text{Gatsby}|c=\text{ALL}|K=10}$, five of the nine pairs are in the set $R_{p=\text{Gatsby}|c=\text{Q\&A}|K=10}$, and two pairs also belong to $R_{p=\text{Gatsby}|c=\text{ALL}|K=5}$. All those seven related discussion pairs contain indeed related discussion posts. By increasing the value of K ($K = 5$ to $K = 10$) and setting $p = \text{Gatsby}|c = \text{ALL}$, the RD-Detector detected two new pairs of related discussion candidates. The maintainer endorsed the relatedness between the discussion posts of the two new pairs.

Homebrew project: Analyzing the sets of related discussion candidates created considering the configuration groups $p = \text{Homebrew}|c = \text{ALL}|K = 5$ and $p = \text{Homebrew}|c = \text{ALL}|K = 10$, the RD-Detector detected 14 new related discussion candidates when changing the K value from 5 to 10. Out of the new 14 candidates, M_Homebrew judged two pairs as unrelated (Table 1). The precision rate decreased from 95%, for $p = \text{Homebrew}|c = \text{ALL}|K = 5$, to 91.17% for $p = \text{Homebrew}|c = \text{ALL}|K = 10$.

Next.js project: The RD-Detector detected 43 new related discussion candidates when we varied the K value and fixed $p = \text{Next.js}|c = \text{Q\&A}$. Out of the 43 new candidates, SE researchers judged 11 pairs as unrelated. This result means that the approach detected 32 new pairs of related discussions by increasing the value of K . The false positives decreased the RD-Detector precision rate from almost 94% to nearly 89%. This scenario repeats to $p = \text{Next.js}|c = \text{Ideas}$ and $p = \text{Next.js}|c = \text{ALL}$. When we changed the K value from 5 to 10, the approach detected 61 and 89 new related discussion candidates for $c = \text{Ideas}$ and $c = \text{ALL}$, respectively. In total, six and 14 new candidates for $c = \text{Ideas}$ and $c = \text{ALL}$ were judged unrelated by SE researchers, respectively. In both cases, the precision rate decreased.

Because of the intersection relationship between sets of related discussion candidates, SE researchers had already judged 208 of the 220 pairs of related discussion candidates in set $R_{p=\text{Next.js}|c=\text{ALL}|K=5}$. Researchers evaluated the new as being related. Among the new 12 pairs, ten pairs had discussions created by the same user, ten pairs had one of the discussions created as Q&A and the other as Ideas, and one pair had one of the discussions created as Q&A and the other as show-and-tell. Finally, the last pair had one of the discussions created as Ideas and the other as show-and-tell. This finding corroborates the Gatsby project findings. Users create related discussion posts in different categories. Regarding $R_{p=\text{Next.js}|c=\text{ALL}|K=10}$, researchers had already judged 303 out of the 309 related discussion pairs in the set. They also judged two new pairs as being unrelated.

Maintainers can set the value of K according to their respective interests. Decreasing the K value increases the RD-Detector precision rate. Higher precision values ensure greater assertiveness in detecting true positives. Conversely, increasing the K value may reduce the precision rate. Increasing the K value may also increase the number of detected related discussion candidates. Thus, maintainers can choose between a more conservatory (better precision) or a more exploratory approach.

False-positive RD-Detector predictions

Four authors of this paper manually analyzed the false positives presented in Table 1. Based on evidence extracted from the discussion posts, we identified some limitations of the proposed approach. We describe

101 the false positive as follows.

102 *Gatsby project:* The two false-positive cases are related to discussions classified as *Ideas*. The
103 researchers identified that the `RD-Detector` did not capture the project-related specifics of both pairs.
104 Although the two posts of pair #1 (Table 1) address the same topic ('JavaScript linting utility ESLint') and
105 have keywords intersection, they address different problems. The discussion posts of the #2 pair (Table 1)
106 address the topic 'Gatsby GraphQL' and have project keywords intersection. However, the specificity of
107 the issues described in the discussion posts differs.

108 We observed that the approach identified posts with similar topics in both cases. However, it did
109 not identify the project issue specificity. Based on these findings, we observed that the `RD-Detector`
110 could fail to treat particular contexts of software projects. We will call this limitation the 'project-specific
111 limitation.'

112 *Homebrew project:* The false-positive pair of related discussion candidate #3 (Table 1) shows that
113 the strategy captured a high level of abstraction from the two discussions, 'problems with installing
114 Homebrew with Rosetta.' However, the `RD-Detector` did not capture the specificity of the problem;
115 the problem is different. Therefore, the relationship between the two is not confirmed.

116 The analysis of the second *Homebrew*'s unrelated pair, #4 (Table 1), shows that the target discussion
117 contains a link to the master discussion, '... for Homebrew mirror configurations. #1906'. Link references
118 can endorse or refute relationships between discussion posts.

119 In this case, the text fragment containing the link reference does not refute that the discussions are
120 related; however, it does not clearly emphasize that they are related. We analyzed the content of discussion
121 pair #4. We identified a limitation regarding the concept of 'related discussions' that can directly influence
122 the evaluation of the approach. We concluded that interpreting the 'related discussions' concept depends
123 on the evaluators' perspective. We call this limitation 'concept imprecision.'

124 However, M.Homebrew's feedback pointed out that pair #4 is unrelated due to the 'project-specific
125 limitation.' The maintainer claims that both discussions address the same project feature but differ on the
126 issue's specificity. According to M.Homebrew maintainer, one discussion '...is asking what the policy is'
127 and the other one '...is announcing support for a new feature.' We also identified that the 'project-specific
128 limitation' justifies the other two false positives detected for the *Homebrew* project.

129 We also analyzed the discussion content of the unrelated pair #5 (Table 1). Like pair #3, the
130 discussions in #5 address issues related to 'Homebrew installation,' but the `RD-Detector` did not
131 capture the specificity of the problem; however, both discussion posts address the same issue and have
132 similar project-related keywords.

133 *Next.js project:* Analyzing the `Next.js` false-positives, we note that the discussion creators (1) used
134 screenshots to detail or describe the issues and (2) used error logs descriptions to show the stack trace
135 of where the error took place. We also note the predominance of (3) template keywords and (4) project
136 keywords in the false-positive discussion posts.

137 Users can add screenshots in the discussion body to help explain their problems. However, the
138 `RD-Detector` measures the semantic textual similarity between discussions' content. The approach
139 does not use images as a source of evidence. Discussion pairs #8, #11, #16, #24, #26, and #30 (Table 1)
140 exemplify this scenario. After preprocessing, the project' keywords may stand out against the actual
141 discussion content. We also identified the predominance of the template keywords in the false-positive
142 pair #16.

143 Discussions creators also use error logs or descriptions to describe the system's discrepancies or
144 non-conformities. During the preprocessing phase, we remove error descriptions embedded in HTML
145 tags. However, when users use error log content to express their questions, they usually intend to ask for
146 help in solving a specific problem highlighted in the error log content. Removing the log also eliminates
147 the problem specificity. The false-positive discussion pairs #7, #19, #27 - #29, #31, and #33 (Table 1)
148 exemplify the use of error log descriptions in discussion posts.

149 The analysis of pairs #15, #18, #22, and #23 shows that (again) both discussions use the same
150 description templates and have identical project keywords. Finally, the false-positive pairs #6, #9, #10,
151 #12 - #14, #17, #20, #21, #25, and #32 (Table 1) present the same set of keywords, which determine the
152 scope of the project. Keywords match lead to the 'project-specific limitation.'

153 We can propose improvements to the proposed approach based on the maintainers' feedback. For
154 example, we can use the maintainers' judgments to optimize the classifier by providing related and
155 non-related discussion samples. Furthermore, we can design strategies to minimize project-specific

156 limitations by treating the predominance of project and template keywords.

Table 1. False-positive related discussion candidates.

#	ID Master	Title Master	ID Target	Title Target	Similarity value
Gatsby					
1	29766	I want to lint ts file with Gatsby's native support.	32122	On creating a better method to extend the default ESLint configuration	0.8970
2	34105	Add gatsby cli for graphiql only	31205	[docs][guides] better document Gatsby GraphQL APIs	0.8295
Homebrew					
3	153	How to "Rerun the Homebrew installer under Rosetta 2.?"	260	Installing Brew with Rosetta2 fails	0.8733
4	1906	Questions on Homebrew's third-party mirroring policy	1917	Setting up mirrors for Homebrew bottles	0.8528
5	3136	Can't install HomeBrew on Mac M1	384	M1 Mac Homebrew Installing Error	0.8563
Next.js					
6	13617	NextJS not showing TypeScript errors on Runtime	24996	I am unable to setup "typescript-is" with nextjs.	0.8646
7	14416	deploy to Vercel error(Build error occurred)	21107	Deploy Vercel Error 'Error: Command "yarn run build" exited with 1'	0.8522
8	15780	ModuleNotFoundError: Module not found: Error: Can't resolve 'fs'	19154	import sub module of other module is fail	0.8887
9	17205	Head/NextScript cannot be used as a JSX component	36547	TypeScript error when building project with self-made package as project reference	0.8519
10	20361	nextjs with next-routes upgrade	13368	Can I replace next-routes with the new versions?	0.9281
11	23469	API source maps	32920	@sentry/nextjs withSentryConfig enables productionBrowserSourceMaps, in the end, assets sourcemap is open to the public.	0.8541
12	24900	ESLint in Next.js and Create Next App	17320	[RFC] ESLint out of the box	0.8479
13	24996	I am unable to setup "typescript-is" with nextjs.	23871	Typescript - Can you change where Next.js outputs the "next-env.d.ts" file?	0.8593
14	25845	NextJS with Tailwind does not work when importing from Global.css	23195	Tailwind CSS not being bundled in static export	0.8624
15	31831	[i18n] Allow generation of routes with lowercase locales	18834	[i18n] Allow having no default locale for international sites where all languages are equal	0.8348
16	32140	next/image feature request: generate specific sizes for each image	24464	[next/image] - Allow to override global device and image sizes with extra component props	0.8201
17	32243	Defer getServerSideProps on client-side navigation	23921	Allow bypass "getData" of page component on client side navigation	0.8147
18	32344	Automatic locale detection can't be disabled in combination with next export	18834	[i18n] Allow having no default locale for international sites where all languages are equal	0.8275
19	33101	Ignore typescript typing error when building from Next.js v12.0.7	34057	tsconfig.json is getting overwritten if no include is specified	0.8539
20	33261	Nextjs keep redirecting with double locale	30693	[Internationalized Routing] user language wont redirect to locale with country	0.8516
21	33374	Problem with middleware and custom-server	33440	Middleware problem with custom express server	0.9047
22	34387	Built-in CSS support is disabled with no custom CSS configuration	21554	Custom webpack configuration in next.config.js not invoked	0.8685
23	34387	Built-in CSS support is disabled with no custom CSS configuration	25041	Webpack 5 build fails with error: configuration[0].module.rules[3].issuer has an unknown property 'test'	0.8514
24	35369	Invalid hook call when creating npm package using tsdx	26812	Invalid hook call error in Next.js v10 with React v17	0.8574
25	35969	next.config.ts	26430	Create an extendable TSConfig base	0.8375
26	36093	[Question] Couldn't access the app after executing "npm run dev".	19029	Issues completing the "tutorial" from a Windows machine	0.8505
27	36414	Cannot import typescript module with generics from outside the project	16203	Next.js - module not found when importing TypeScript module	0.8585
28	37505	Deployment failed with error-Failed to compile.	16915	Deploying Nextjs app failed on versel	0.8957
29	37505	Deployment failed with error-Failed to compile.	26643	[Solved] Error deploying Next.js app on Vercel	0.8531
30	38829	Next.js Dynamic Routes and getServersSideProps occur error while export	25744	Build Error When Using getStaticProps	0.8675
31	39205	Vercel building failed.	20778	Vercel deployment failed - Nextjs Blog Example	0.9000
32	39719	Deactivate "trailingSlash" logic	27688	next.config.js redirects and trailingSlash - allow redirects to be priority	0.8263
33	40456	Got error when starting nextjs project	39501	NextJS getting error on the first npm run dev	0.8669

157 ACKNOWLEDGMENTS

158 We would like to thank GitHub and the OSS maintainers for supporting this research. We also would like to
 159 thank the financial support granted by CNPq through processes number 314174/ 2020-6 and 313067/2020-
 160 1. CAPES financial code 001. FAPESP under grant 2020/05191-2. FAPEAM through process number
 161 062.00150/2020. This research was also carried out within the scope of the Samsung-UFAM Project for
 162 Education and Research (SUPER), according to Article 48 of Decree number 6.008/2006(SUFRAMA).