Towards Understanding Emotions in Informal Developer Interactions: A Gitter Chat Study

Amirali Sajadi
Drexel University
Philadelphia, PA, USA
amirali.sajadi@drexel.edu

Kostadin Damevski
Virginia Commonwealth University
Richmond, Virginia, USA
kdamevski@vcu.edu

Preetha Chatterjee
Drexel University
Philadelphia, PA, USA
preetha.chatterjee@drexel.edu

ABSTRACT

Emotions play a significant role in teamwork and collaborative activities like software development. While researchers have analyzed developer emotions in various software artifacts (e.g., issues, pull requests), few studies have focused on understanding the broad spectrum of emotions expressed in chats. As one of the most widely used means of communication, chats contain valuable information in the form of informal conversations, such as negative perspectives about adopting a tool. In this paper, we present a dataset of developer chat messages manually annotated with a wide range of emotion labels (and sub-labels), and analyze the type of information present in those messages. We also investigate the unique signals of emotions specific to chats and distinguish them from other forms of software communication. Our findings suggest that chats have fewer expressions of *Approval* and *Fear* but more expressions of *Curiosity* compared to GitHub comments. We also notice that *Confusion* is frequently observed when discussing programming-related information such as unexpected software behavior. Overall, our study highlights the potential of mining emotions in developer chats for supporting software maintenance and evolution tools.

CCS CONCEPTS

• **Software and its engineering → Collaboration in software development**.

KEYWORDS

emotion analysis, software developer chats

1 INTRODUCTION

Emotions can greatly influence teamwork and collaborative activities such as software development. Specific tasks have been

found to be significantly impacted by developer emotions, e.g., bug fixing [13, 28], and build success of continuous integration [35]. Researchers have extensively studied how developer emotions affect software development, created approaches for automatically detecting emotions [4, 10, 21, 22] and, in cases, provided recommendations for the developers [7, 14, 17]. More recently, researchers have studied complex, emotionally charged psychological concepts such as toxicity in issue reports [16, 26], and confusion in code reviews [14]. Likewise affective trust between developers of a project was investigated in pull requests and commit comments [3, 11, 31, 32].

A significant amount of research has been conducted on analyzing the emotions of developers in various software artifacts, such as issues, and pull requests [3, 20, 27]. However, surprisingly, there is a lack of studies on understanding emotions on chat platforms, despite their widespread use among software developers. Other aspects of developer chats have been previously studied and it was shown that chats are generally interactive and often used for informal communications [6, 34] which intuitively makes these communication platforms a suitable place to express emotions. Chatterjee et al. noticed that expression of developer emotions is prevalent in chat communications on platforms such as Slack, IRC, and Discord [8, 37]. Kuutila et al. investigated Slack and Hipchat to analyze developers' sentiments and their impact on productivity [23]. However, none of these studies systematically analyze developer emotions in textual chat messages.

In this paper we investigate different types of emotions expressed in developers' chat communications. We first select a subset of 400 developer chat messages from the Gittercom dataset [30], and then manually annotate it with emotion categories (e.g., *Anger*, *Joy*) using Imran et al's extended emotion taxonomy [20]. Additionally, we leveraged Pan et al.'s taxonomy to determine the type of information shared (e.g., programming problems, task progress) in these messages. Next, we qualitatively analyze the dataset of 400 messages to understand the relationship between the type of information conveyed and the type of emotion expressed in the messages. We aim to answer the following research questions: **(RQ1:)** What types of emotions are expressed in developer chats and how are they associated with specific types of information or developer intent?; **(RQ2:)** How do emotions expressed in chats differ from emotions expressed in other forms of software communications? What are the specific signals of emotions that are unique to chats?

Our findings show that, compared to the GitHub issue or pull request comments, chat messages contain fewer instances of expressing *Approval* and *Fear* and more instances of expressing *Curiosity*, which is expected since chat communications are informal in nature and often follow a Q&A format. Chat communications also consists of more emoticons, shorter sentences, and a generally

more informal tone. These observations can improve the effectiveness of automatic emotion detection tools by providing insights into the modifications they require for adapting to chat platforms.

2 METHODOLOGY

Emotion Categories and Detection Tool. Shaver's taxonomy, widely used in various software engineering studies [5, 20, 27], is a hierarchical, tree-structured emotion representation model, consisting of three levels. The top level comprises six basic emotions: *Anger*, *Love*, *Fear*, *Joy*, *Sadness*, and *Surprise*. For each basic emotion, there exist secondary and tertiary-level emotions that provide more refined granularity for the preceding level. For instance, *Optimism* and *Hope* are the secondary and tertiary level emotions, respectively, for *Joy*. In a recent work, Imran et al. noticed that certain emotions commonly expressed in developer communications were absent from Shaver's framework [20]. Therefore, they extended Shaver's categories with select emotions from GoEmotions [12]. In this study, we use Imran et al's extended taxonomy (Table 1).

Data Selection. For this study, we use GitterCom, a dataset, consisting of 10,000 messages collected from 10 Gitter communities [30]. In order to obtain a goldset, we selected a subset of 400 messages to be manually annotated with corresponding emotion labels. In order to obtain a statistically significant sample with confidence of 95%±5%, we sampled 400 messages distributing the samples equally across 4 different communities [2]. As a measure to avoid the inclusion of text that does not exhibit any emotion, we have decided to limit our sampling to the instances that contain either a positive or negative sentiment, since intuitively messages with a stronger sentiment have a higher potential for expressing emotion. To pre-process the data, we removed stopwords, urls, and user mentions from the messages. We also converted the text to lower-case, and tokenized the words. NLTK VADER [19] was used to automatically assign a value between -1 and 1 to each message, representing its sentiment. The authors then randomly selected 200 instances from the messages with sentiment scores in the first quartile of the value distribution (strongly positive) and 200 instances from the messages with sentiment values in the last quartile of the value distribution (strongly negative). These 400 messages contain 100 chat messages from each of the four GitterCom projects with the highest number of users (i.e., *scikit-learn*, *Marionette*, *jHipster*, and *UIkit*).

Dataset Annotation. Two human judges with (3+ years) experience in programming and familiarity with the Gitter platform annotated the 400 selected messages. The annotators were given instructions, similar to ones presented in Imran et al.'s study [20], containing details on emotion categories, subcategories, definitions and examples. The judges were asked to determine whether each message expresses any of the six basic emotions along with their secondary and tertiary subcategories. Since chat messages in the GitterCom dataset are, in many cases, not a single sentence, one can often extract the context for each message exhibiting an emotion. The annotations, therefore, relied on the emotion expressed through the entire chat message rather than single sentences.

Next, we adopted Pan et al.'s taxonomy to determine the types of information available in the 400 messages [29]. This taxonomy categorizes developers' chat communications into the following categories: (1) Problem Report: Conversation regarding unexpected

Table 1: Extended Taxonomy of Shaver's Tree-structured Emotion Categories [20]

Basic Emotion	Secondary Emotion	Tertiary Emotion
	Irritation	Annoyance, Agitation, Grumpiness, Aggravation, Grouchiness
	Exasperation	Frustration
	Rage	Anger, Fury, Hate, Dislike, Resentment, Outrage, Wrath, Hostility, Bitterness, Ferocity, Loathing, Scorn, Spite, Vengefulness
Anger	Envy	Jealousy
	Disgust	Revulsion, Contempt, Loathing
	Torment	-
	Disapproval	-
	Affection	Liking, Caring, Compassion, Fondness, Affection, Love, Attraction, Tenderness, Sentimentality, Adoration
Love	Lust	Desire, Passion, Infatuation
	Longing	-
	Horror	Alarm, Fright, Panic, Terror, Fear, Hysteria, Shock, Mortification
Fear	Nervousness	Anxiety, Distress, Worry, Uneasiness, Tenseness, Apprehension, Dread
	Cheerfulness	Happiness, Amusement, Satisfaction, Bliss, Gaiety, Glee, Jolliness, Joviality, Joy, Delight, Enjoyment, Gladness, Jubilation, Elation, Ecstasy, Euphoria
	Zest	Enthusiasm, Excitement, Thrill, Zeal, Exhilaration
	Contentment	Pleasure
	Optimism	Eagerness, Hope
Joy	Pride	Triumph
	Enthrallment	Enthrallment, Rapture
	Relief	-
	Approval	-
	Admiration	-
	Suffering	Hurt, Anguish, Agony
	Sadness	Depression, Sorrow, Despair, Gloom, Hopelessness, Glumness, Unhappiness, Grief, Woe, Misery, Melancholy
	Disappoint	Displeasure, Dismay
Sadness	Shame	Guilt, Regret, Remorse
	Neglect	Embarrassment, Insecurity, Insult, Rejection, Alienation, Isolation, Loneliness, Homesickness, Defeat, Dejection, Humiliation
	Sympathy	Pity
	Surprise	Amazement, Astonishment
	Confusion	-
Surprise	Curiosity	-
	Realization	-

behaviors or bug reports, containing information about (a) *Programming problems,* (b) *Library Problems*, or (c) *Documentation problems*; (2) Information Retrieval: Conversations initiated and carried on with the purpose of acquiring or providing information about a certain topic such as: (a) *Programming information*, (b) *Library information*, (c) *Documentation information*, or (d) *General information* e.g., choice of technology; (3) Project Management: Discussion among contributors and team members about the overall state of their project and the future plans for their work, such as: (a)*Technical discussion* or (b) *Task progress* e.g., release schedules.

Following the initial annotation phase, the inter-rater agreement for each emotion category was calculated using Cohen's Kappa. The resulting values were substantial for *Joy* and *Love* (above 0.6) and moderate for the remaining four emotions (ranging from 0.41 to 0.60) [36]. To ensure the best possible results, the annotators held multiple discussions to resolve their disagreements and reevaluated their annotations iteratively. This process continued until they reached a Cohen's Kappa value of 1 and resolved all disagreements.

3 PRELIMINARY RESULTS AND DISCUSSION

RQ1. What types of emotions are expressed in developer chats and how are they associated with specific types of information or developer intent?

Table 2: Example Messages from GitterCom, GitHub Comments, and their Corresponding Annotated Emotions.

Emotion – *Primary (Secondary, Tertiary)*	Chat Messages	GitHub Comments (Issues and Pull Requests)
Anger *(Exasperation, Frustration)*	I can be patient for first time but for each prod build it checks again and thats annoying...	Too bad! Thank you anyway ... This issue is really driving me nuts ...
Fear *(Nervousness, Anxiety)*	I always get an error telling me my request can't be processed :worried:	I can't check this locally because they fail with same error for me even on master
Joy *(Cheerfulness, Satisfaction)*	Yea, I'm continually impressed by the community and diversity of the conversation, prs, issues, etc...	You are 100% on the money with this. Turns out the parsing was incorrect
Sadness *(Sadness, Unhappiness)*	@ogrisel what is your plan for the day? I didn't have much time on the weekend unfortunately :-/	I don't think I can, since it is an implementing class. The analyzer is unhappy with it.
Love *(Affection, Fondness)*	Thank you and i really appreciate your time on this wonderful framework, I love using it :)	PS: I am fan of yours, I love your content out there! :smiley:
Surprise *(Confusion, Amazement)*	it did? I didn't see that. haha I know the open tabs issue. Well the "linear" broke some cases of "fit" and "fit_transform" not doing the same thing. Maybe it broke other things, too.	"true" feels like magic. Maybe it should be a default value provided to the set or a symbol
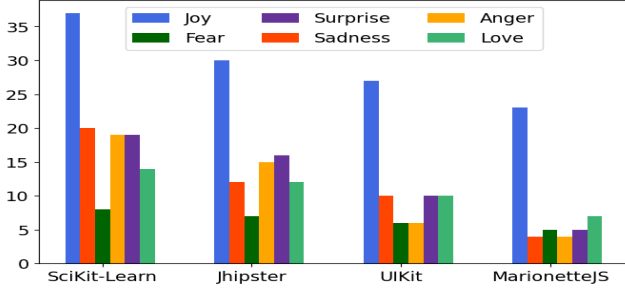
Figure 1: Project-wise Freq. of Messages Exhibiting Emotion.

264 out of 400 messages (66%) contained at least one emotion, while 136 messages (34%) expressed no emotions. As indicated by Figure 1, *Joy* is the most prevalent emotion expressed in our dataset. Within the 264 messages that contain emotions, 117 (44.3%) exhibit *Joy*, 50 (18.9%) *Surprise*, 46 (17.4%) *Sadness*, 44 (16.6%) *Anger*, 43 (16.2%) *Love*, and 26 (9.8%) *Fear*. Fig 1 illustrates the distribution of the emotions in our dataset across the four Gitter projects. *Joy* is consistently the most frequent emotion across all projects while *Fear* stays the least expressed emotion in three of the projects. Overall *Scikit-Learn*, *Jhipster*, and *UIKit* have rather similar distributions of emotions. MarionetteJS, however, tends to be different since it has a more even distribution over the six emotions. In Figure 2 we show the results of the annotations for the first, second, and third-level emotions. Overall, *Joy* and *Sadness* stem from more diverse secondary and tertiary categories compared to the other four basic emotions. *Joy* is dominated by the second-level emotions of *Cheerfulness* and *Zest*, which can be an indicator of a positive attitude and environment in the communities we explored. *Sadness*, on the other hand, is more evenly split across two of its second-level categories: *Sadness*, which generally expresses a level of dissatisfaction towards the topic, and *Disappointment*, which has often been directed towards unexpected behaviors in software. *Anger* and *Love*, in contrast, are more consistent across the emotion levels. *Love*, for example, almost always is categorized as *Affection* and *Fondness* on the second and third levels.

The annotations of the information categories show that the four most common types of information available in the developers'
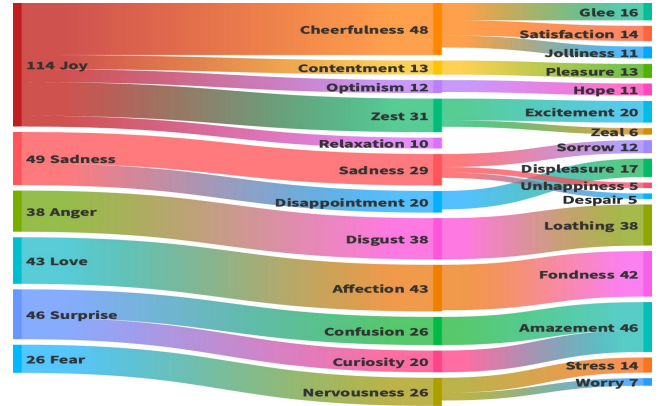
Figure 2: Distribution of the Base, Second, and Third-Level Emotions (n >= 5).

chats are *Technical progress*, *Programming information*, *Technical discussion*, and *Programming problem* with 31, 30, 28, and 24 instances, respectively. The categories related to *Library* and *Documentation* were scarce, with less than 10 occurrences for each category. These results highlight the prevalence of the exchange of information regarding general project management, programming-related problem reports, and programming-related information retrieval in our dataset. Furthermore, we observed a consistent trend of messages discussing the technical progress of a project being accompanied by positive emotions, such as *Joy* and *Love*. The messages containing the *Surprise* emotion, are correlated with those inquiring information about programming or those reporting programming-related problems. We also noticed that chat conversations frequently involve discussions about unexpected software behavior, which can elicit *Confusion*, a sub-category of *Surprise* in our taxonomy.

RQ2. How do emotions expressed in chats differ from emotions expressed in other forms of software communications? What are the specific signals of emotions that are unique to chats?

To identify the unique characteristics of developers' chat communications, we compare our emotion annotations on chat messages with the emotions in Imran et al.'s dataset of GitHub issues and pull request comments with either positive or negative sentiment [20].

Table 2 contains examples of chat messages (from our dataset), along with examples of GitHub comments (from Imran et al.'s dataset) exhibiting the same emotions. In general, chat messages tend to facilitate more informal conversations, in keeping with the nature of the communication tool. As one would expect, the length of the messages in chat communications are generally shorter. GitterCom dataset messages are on average 5.87 words long (5.09 in our sampled 400 instances), while the 2000 GitHub comments in Imran et al.'s dataset are on average 12.82 words, i.e., more than twice as long. We also notice a pattern in the more frequent use of emoticons in chat messages compared to GitHub comments. Emoticons are often indicators of implicit emotions, and understanding them could potentially improve the performance of existing emotion detection tools on software engineering-specific text [10].

We observe that a significant number of instances labeled as *Joy* in the GitHub data (~18%) are exhibiting *Approval* on the second level. In contrast, in our Gitter dataset only two instances out of the 117 chat messages that exhibit *Joy* express the writer's approval. This inconsistency is expected since GitHub communications often entail the evaluation of one's contribution to the project. For instance, the changes suggested through a pull request in many repositories are required to be approved by at least one reviewer prior to being merged. Therefore, approvals exhibiting positive emotions such as *Joy* are more frequent in GitHub.

Developers' chat communications predominantly follow Q&A formats [9], which may explain the higher prevalence of instances labeled as *Curiosity* in our dataset. Around 3.5% of the GitHub issue and pull request comments presented in Imran et al.'s dataset exhibit *Curiosity*, whereas 6.5% of the Gitter chat communications in our dataset demonstrate some form of this emotion. GitHub comments may elicit more negative emotions (e.g., *Fear* and by extension its subcategories, *Nervousness*, *Worry*, and *Stress*), since they are often used to discuss issues or bugs in the code and negative emotions such as *Fear* are commonly associated with uncertainty and risk. In line with these expectations, we observe that *Fear* was present in 9.9% of the GitHub comments and 6.5% of the chat messages. Imran et al.'s annotations also point to some instances of *Fear* containing *Horror*, a 2nd level emotion absent in the our chat messages.

4 IMPLICATIONS

To the best of our knowledge, this study took the first step toward systematically analyzing emotions in developer chat communications. Our analysis of the Gitter dataset revealed a range of emotions expressed by developers on chats, predominantly *Joy*, *Surprise*, and *Sadness*. We noticed that technical progress in software development often evokes positive emotions such as *Joy* and *Love*. In contrast, unexpected software behavior or bugs tends to elicit negative emotions such as *Sadness* and *Anger*. These findings emphasize the potential to develop automated interventions, such as emotion-detection bots, that take into account users' emotional responses to unexpected events, and suggest potential solutions in a timely manner [15, 18, 25]. In addition to the basic emotions, the secondary and tertiary-level emotions enables us to further analyze the messages and can shed light on the underlying causes of the dominant emotions in informal developer conversations. For instance, *Amazement* or *Confusion*, subcategories of *Surprise*, were

exhibited in chat messages detailing the satisfactory or unexpected performance of a newly adopted tool. The dominant presence of *Curiosity* in chat communications compared to the GitHub comments confirms the prevalence of the Q&A questions in chats. This suggests that chats are better mining source to design Q&A-based systems such as conversational search assistants [7, 34].

Limited availability of ground truth data has hindered the extensive evaluation of existing approaches for emotion detection in software developers' written text [24]. As a first step towards addressing this challenge, we present a dataset of 400 developer chat messages annotated with Imran et al's extended emotion taxonomy [20], originally based on Shaver's taxonomy [33]. Our dataset can be leveraged to analyze developer emotions across various channels (e.g., chats, pull requests) in a software project. Training tools on different communication channels in a project, including chats, offers the potential for building project-specific emotion detectors.

Overall, tools and applications that aim to improve software development processes and team communication can benefit from mining developer chats. Chats can provide rich contextual information on how developers interact and collaborate in real-time, which can help identify communication gaps and improve team dynamics. Proactively identifying negative emotions expressed in chat conversations can help detect potential conflicts, prevent burnout, and improve team collaboration. By analyzing the emotions expressed by developers towards specific aspects of a project, one can assess their opinions on particular tools and technologies. Compared to other artifacts such as issue comments, chats can offer a more informal and nuanced perspective on developer emotions and interactions. Chats are often more conversational and spontaneous, allowing for a broader range of emotions and expressions (e.g., burnout) that may not be captured in other types of communication.

5 CONCLUSION AND FUTURE WORK

The increasing reliance on chat platforms for virtual communication among OSS teams and developers in general highlights the significance of studying these communications to gain a better understanding of the development process. As a distinct type of messaging tool, chat platforms provide a unique form of communication that allows developers to express themselves more spontaneously, including exhibiting emotions about specific aspects of a project. We present a dataset of 400 messages from Gitter, manually annotated to identify prevalent emotions in developer chats. Our findings shed light on how emotions are expressed on chat platforms and how they differ from emotions expressed in GitHub issue and pull requests. We also explored how the emotions vary based on developer intent and the type of information exchanged in the chat messages. Our immediate next steps focus on expanding to a larger dataset, manually annotated with the emotion categories. This step will help us establish, with confidence, the areas in which the current automatic emotion detection tools are lacking (e.g., certain emotions or emotion subcategories) and lay the foundation on which we can further develop and improve these tools. We make our annotated dataset publicly available for future research [1].

REFERENCES

[1] 2023. dataset:. https://figshare.com/articles/dataset/Towards_Understanding_Emotions_in_Informal_Developer_Interactions_A_Gitter_Chat_Study/23796138. Publication date unknown.

[2] Mohamad Adam Bujang and Nurakmal Baharum. 2017. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. *Archives of Orofacial Science* 12, 1 (2017).

[3] Fabio Calefato and Filippo Lanubile. 2016. Affective trust as a predictor of successful collaboration in distributed software projects. In *2016 IEEE/ACM 1st International Workshop on Emotional Awareness in Software Engineering (SEmotion)*. IEEE, 3–5.

[4] Fabio Calefato, Filippo Lanubile, and Nicole Novielli. 2017. Emotxt: a toolkit for emotion recognition from text. In *2017 seventh international conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 79–80.

[5] Fabio Calefato, Filippo Lanubile, Nicole Novielli, and Luigi Quaranta. [n. d.]. EMTk - The Emotion Mining Toolkit. In *2019 IEEE/ACM 4th International Workshop on Emotion Awareness in Software Engineering (SEmotion)*. https://doi.org/10.1109/SEmotion.2019.00014

[6] Preetha Chatterjee, Kostadin Damevski, Nicholas A Kraft, and Lori Pollock. 2021. Automatically identifying the quality of developer chats for post hoc use. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 30, 4 (2021), 1–28.

[7] Preetha Chatterjee, Kostadin Damevski, and Lori Pollock. 2021. Automatic Extraction of Opinion-based Q&A from Online Developer Chats. In *Proceedings of the 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. 1260–1272. https://doi.org/10.1109/ICSE43902.2021.00115

[8] P. Chatterjee, K. Damevski, L. Pollock, V. Augustine, and N.A. Kraft. 2019. Exploratory Study of Slack Q&A Chats as a Mining Source for Software Engineering Tools. In *Proceedings of the 16th International Conference on Mining Software Repositories (MSR'19)* (Montreal, Canada). https://doi.org/10.1109/MSR.2019.00075

[9] Preetha Chatterjee, Kostadin Damevski, Lori Pollock, Vinay Augustine, and Nicholas A Kraft. 2019. Exploratory study of slack q&a chats as a mining source for software engineering tools. In *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*. IEEE, 490–501.

[10] Zhenpeng Chen, Yanbin Cao, Xuan Lu, Qiaozhu Mei, and Xuanzhe Liu. 2019. Sentimoji: an emoji-powered learning approach for sentiment analysis in software engineering. In *Proceedings of the 2019 27th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*. 841–852.

[11] Guilherme Augusto Maldonado da Cruz, Elisa Hatsue Moriya-Huzita, and Valéria Delisandra Feltrim. 2018. ARSENAL-GSD: A framework for trust estimation in virtual teams based on sentiment analysis. *Information and Software Technology* 95 (2018), 46–61. https://doi.org/10.1016/j.infsof.2017.10.016

[12] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan S. Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A Dataset of Fine-Grained Emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, 4040–4054. https://doi.org/10.18653/v1/2020.acl-main.372

[13] Giuseppe Destefanis, Marco Ortu, S. Counsell, S. Swift, M. Marchesi, and R. Tonelli. 2016. Software development: do good manners matter? *PeerJ Comput. Sci.* 2 (2016), e73.

[14] Felipe Ebert, Fernando Castor, Nicole Novielli, and Alexander Serebrenik. 2021. An exploratory study on confusion in code reviews. *Empirical Software Engineering* 26 (2021), 1–48.

[15] Linda Erlenhov, Francisco Gomes de Oliveira Neto, and Philipp Leitner. 2020. An Empirical Study of Bots in Software Development: Characteristics and Challenges from a Practitioner's Perspective. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Virtual Event, USA) *(ESEC/FSE 2020)*. Association for Computing Machinery, New York, NY, USA, 445–455. https://doi.org/10.1145/3368089.3409680

[16] Isabella Ferreira, Bram Adams, and Jinghui Cheng. 2022. How heated is it? Understanding GitHub locked issues. (2022).

[17] G. Fucci, N. Cassee, F. Zampetti, N. Novielli, A. Serebrenik, and M. Di Penta. 2021. Waiting Around or job Half-Done? Sentiment in Self-Admitted Technical Debt. In *2021 2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR) (MSR)*. IEEE Computer Society, Los Alamitos, CA, USA, 403–414. https://doi.org/10.1109/MSR52588.2021.00052

[18] Anze Gao, Sihao Chen, Tao Wang, and Jinsheng Deng. 2022. Understanding the Impact of Bots on Developers Sentiment and Project Progress. In *2022 IEEE 13th International Conference on Software Engineering and Service Science (ICSESS)*. 93–96. https://doi.org/10.1109/ICSESS54813.2022.9930282

[19] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.

[20] Mia Mohammad Imran, Yashasvi Jain, Preetha Chatterjee, and Kostadin Damevski. 2023. Data Augmentation for Improving Emotion Recognition in Software Engineering Communication. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering* (Rochester, MI, USA) *(ASE '22)*. Association for Computing Machinery, New York, NY, USA, Article 29, 13 pages. https://doi.org/10.1145/3551349.3556925

[21] Md Rakibul Islam, Md Kauser Ahmmed, and Minhaz F. Zibran. 2019. MarValous: Machine Learning Based Detection of Emotions in the Valence-Arousal Space in Software Engineering Text. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* (Limassol, Cyprus) *(SAC '19)*. Association for Computing Machinery, New York, NY, USA, 1786–1793. https://doi.org/10.1145/3297280.3297455

[22] Md Rakibul Islam and Minhaz F Zibran. 2018. DEVA: sensing emotions in the valence arousal space in software engineering text. In *Proceedings of the 33rd annual ACM symposium on applied computing*. 1536–1543.

[23] Miikka Kuutila, M. Mäntylä, and Maëlick Claes. 2020. Chat activity is a better predictor than chat sentiment on software developers productivity. *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops* (2020).

[24] Bin Lin, Nathan Cassee, Alexander Serebrenik, Gabriele Bavota, Nicole Novielli, and Michele Lanza. 2022. Opinion Mining for Software Development: A Systematic Literature Review. *ACM Trans. Softw. Eng. Methodol.*, Article 38 (2022), 41 pages. https://doi.org/10.1145/3490388

[25] Bin Lin, Alexey Zagalsky, Margaret-Anne Storey, and Alexander Serebrenik. 2016. Why Developers Are Slacking Off: Understanding How Software Teams Use Slack. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (San Francisco, California, USA) *(CSCW '16 Companion)*. ACM, New York, NY, USA, 333–336. https://doi.org/10.1145/2818052.2869117

[26] Courtney Miller, Sophie Cohen, Daniel Klug, Bogdan Vasilescu, and Christian Kästner. 2022. "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In *In 44th International Conference on Software Engineering (ICSE'22)*.

[27] Alessandro Murgia, Marco Ortu, Parastou Tourani, Bram Adams, and Serge Demeyer. 2018. An Exploratory Qualitative and Quantitative Analysis of Emotions in Issue Report Comments of Open Source Systems. *Empirical Softw. Engg.* 23, 1 (feb 2018), 521–564. https://doi.org/10.1007/s10664-017-9526-0

[28] Marco Ortu, Alessandro Murgia, Giuseppe Destefanis, Parastou Tourani, Roberto Tonelli, Michele Marchesi, and Bram Adams. [n. d.]. The Emotional Side of Software Developers in JIRA. In *Proceedings of the 13th International Conference on Mining Software Repositories* (Austin, Texas) *(MSR '16)*. https://doi.org/10.1145/2901739.2903505

[29] Shengyi Pan, Lingfeng Bao, Xiaoxue Ren, Xin Xia, David Lo, and Shanping Li. 2021. Automating developer chat mining. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 854–866.

[30] Esteban Parra, Ashley Ellis, and Sonia Haiduc. 2020. Gittercom: A dataset of open source developer communications in gitter. In *Proceedings of the 17th International Conference on Mining Software Repositories*.

[31] Amirali Sajadi, Kostadin Damevski, and Preetha Chatterjee. 2023. Interpersonal Trust in OSS: Exploring Dimensions of Trust in GitHub Pull Requests. In *Proceedings of the 45th International Conference on Software Engineering (NIER Track)* (Melbourne, Australia) *(ICSE '23)*.

[32] Hitesh Sapkota, Pradeep K. Murukannaiah, and Yi Wang. 2020. A network-centric approach for estimating trust between open source software developers. *PLOS ONE* 14, 12 (12 2020), 1–30. https://doi.org/10.1371/journal.pone.0226281

[33] Phillip Shaver, Judith Schwartz, Donald Kirson, and Cary O'connor. 1987. Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology* 52, 6 (1987), 1061.

[34] Lin Shi, Xiao Chen, Ye Yang, Hanzhi Jiang, Ziyou Jiang, Nan Niu, and Qing Wang. 2021. A First Look at Developers' Live Chat on Gitter. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Athens, Greece) *(ESEC/FSE 2021)*. Association for Computing Machinery, New York, NY, USA, 391–403. https://doi.org/10.1145/3468264.3468562

[35] Rodrigo Souza and Bruno Silva. 2017. Sentiment Analysis of Travis CI Builds. In *Proceedings of 14th International Conference on Mining Software Repositories (MSR)*. https://doi.org/10.1109/MSR.2017.27

[36] Steven E Stemler. 2004. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation* 9, 1 (2004), 4.

[37] Keerthana Muthu Subash, Lakshmi Prasanna Kumar, Sri Lakshmi Vadlamani, Preetha Chatterjee, and Olga Baysal. 2022. DISCO: A Dataset of Discord Chat Conversations for Software Engineering Research. In *Proceedings of the 19th International Conference on Mining Software Repositories (MSR'22)*.