

Investigando os desafios em atividades de crowdtesting: um estudo exploratório em plataformas *on-demand matching*

Jessica Bosio¹, Alexandre Zanatta¹, Leticia Santos Machado²

¹Instituto de Ciências Exatas e Geociências – Universidade de Passo Fundo (UPF)

²Instituto de Ciências Exatas e Naturais – Universidade Federal do Pará (UFPA)

jebosio@gmail.com, zanatta@upf.br, leticia.smachado@gmail.com

Abstract. *Crowdsourced Software Testing (CST) is the outsourcing of software testing to an online community in the crowdsourcing model. Crowdtesters play the important role in the CST by being an active member. However, past research has suggested that crowdworkers face many barriers when making their contributions in such communities. The goal of our research was to identify what are the difficulties that block the contribution in CST. To provide a preliminary set of problems we conducted a qualitative study. We identified seven preliminary types of barriers that hinder crowdtesters' contribution. For future work we'll provide suggestions on how to minimize the barriers.*

Resumo. *Crowdsourced Software Testing (CST) é a terceirização de teste de software para uma comunidade on-line no modelo crowdsourcing. Testadores executam um importante papel em CST para se tornarem membros ativos. Entretanto, pesquisas anteriores mostram que trabalhadores na multidão enfrentam muitas dificuldades para contribuir em comunidades on-line. O objetivo do trabalho foi identificar quais são as dificuldades que impedem a contribuição em atividades de CST. Realizou-se um estudo qualitativo para identificar um conjunto preliminar de problemas. Foram identificadas 7 barreiras que dificultam a participação. Para trabalho futuro, serão apresentadas sugestões para minimizar estas barreiras.*

1. Introdução

O crowdsourcing e o mercado laboral on-line surgiram como um novo ambiente de trabalho humano permitindo que as organizações dimensionem, de maneira flexível, sua força de trabalho e contratem especialistas independentes, distribuídos geograficamente, com conhecimentos e habilidades específicas conforme demanda.

Este ambiente inclui atividades relacionados ao desenvolvimento de software, entre eles, o teste de software, conhecido como *Crowdsourced Software Testing (CST)* (MAO, LICIA, *et al.*, 2016). CST tem como principal objetivo validar a qualidade de um software por meio da realização de testes pela multidão. Porém, a diversidade advinda de uma grande quantidade de pessoas interagindo para executar atividades e resolver problemas, cria oportunidades, mas também potencializa os desafios relacionados a coordenação e comunicação (STOL e FITZGERALD, 2014).

Ambientes de crowdsourcing, como o CST são altamente dependentes da motivação e participação das pessoas para o seu sucesso pois os participantes “têm uma alta tendência a interromper sua participação depois de algum tempo” como destacam Lu et al. (LU, PHANG e YU, 2011). Segundo Wu et al. (WU, LI e TSA, 2013), apenas 20% dos usuários da Topcoder (plataforma de software crowdsourcing) são participantes ativos, ou seja, participantes que submetem suas soluções para as tarefas de software apresentadas pela plataforma TopCoder. Diferentes fusos horários, distribuição geográfica das pessoas, e a falta de conhecimento da língua nativa da tarefa são apontados por Saremi e Yang

(SAREMI e YANG, 2015) como aspectos motivacionais que influenciam o trabalhador da multidão (*crowd*) a concluir ou não sua tarefa. Neste sentido, Yang *et al.* (YANG, KARIM, *et al.*, 2016) apontam um índice de 82,9% de taxa de desistência de submissão de tarefas de software crowdsourcing, considerando que a maioria dos participantes em comunidades on-line não contribuem, como apontam Hill *et al.* (HILL, HOLLAN, *et al.*, 1992).

Assim, o objetivo deste trabalho é iniciar um estudo exploratório para identificar quais são os fatores que influenciam os trabalhadores da multidão de completarem suas tarefas e reportarem seus casos de teste em CST além disso, investigar quais mecanismos poderiam ser propostos para auxiliá-los a finalizarem estas tarefas. Entende-se, portanto, que existe oportunidade para uma investigação uma vez que a desinteresse em seguir contribuindo, participando e aceitando atividades de testes pode ser um risco ao modelo de CST que precisa da força de trabalho online dos membros das plataformas para atender as demandas de teste de seus clientes. Assim, a questão de pesquisa do presente trabalho é: *“Quais são as barreiras que um trabalhador na multidão enfrenta ao realizar testes funcionais em atividades de CST?”*.

2. Plataformas de CST

O modelo crowdsourcing possui quatro elementos: a multidão, a plataforma, o cliente e a atividade. A seguir far-se-á uma breve descrição das plataformas utilizadas no trabalho.

A uTest¹, considerada a pioneira de CST no Brasil, atua em mais de 150 países com aproximadamente 60 mil testadores. Além do recebimento de uma compensação financeira (remuneração monetária), os testadores participam de um sistema de reconhecimento e reputação e ranking. A plataforma utiliza apenas o idioma inglês e seus testes são do tipo funcionais. Na BugFinders², os testes geralmente não possuem casos de teste ou especificação do escopo do projeto, são considerados testes a nível mais exploratório, o que torna o relato de defeitos um pouco complexo. Essa plataforma também possui o idioma inglês como padrão e seus testes são funcionais. No TestBirds³ o participante não necessita de nenhum conhecimento prévio em testes de software ou tecnologias em particular. O participante também realiza um projeto de teste não remunerado, para avaliação de suas capacidades de encontrar e reportar defeitos, porém, essa plataforma exige o uso de imagens detalhando esses defeitos. O TestBirds também utiliza o idioma inglês e segue nos testes funcionais. No PassBrains⁴ o participante necessita aguardar um convite para integrar um time de testes de projeto remunerado, baseado nos seus conhecimentos e dispositivos informados no cadastro. Seus testes são do tipo funcional e o padrão do idioma é o inglês. O objetivo do 99Tests⁵ é fornecer testes funcionais de segurança, de carga e teste de automação. Também trabalha com processos em ciclos de testes, seguindo exigências quanto aos componentes que devem ser validados em cada ciclo. Essa plataforma também utiliza o idioma inglês.

3. Metodologia

Esse trabalho caracteriza-se como uma pesquisa qualitativa e os dados foram obtidos por meio da realização de testes funcionais em cinco plataformas de CST. Foram eleitas apenas

1 <https://www.utest.com/>

2 <https://join.digivante.com/as-functional-tester>

3 <https://www.testbirds.com/>

4 <https://www.passbrains.com/>

5 <https://99tests.com/>

as plataformas que realizam atividades sob demanda (*on-demand matching*), exclusivamente para testes de software.

Como os registros das atividades de teste realizadas foram armazenados em diários (com formato de livre redação pelos participantes) individuais utilizou-se a Análise de Conteúdo (BARDIN, 1977), para compilar os dados e selecionar aspectos que seguem um padrão de repetição, e assim gerar os códigos, que descrevessem em forma de uma única palavra tais padrões. Posteriormente os códigos foram analisados e categorizados. A ferramenta Nvivo⁶ auxiliou o processo de organização dos dados coletados.

A coleta de dados foi realizada por meio da submissão de tarefas de teste funcional de software nas plataformas, observando-se as dificuldades encontradas em cada etapa durante a realização das ações de identificação e reporte dos casos de teste executados. Essas tarefas referem-se à participação de um trabalhador na multidão, desde o cadastro nas plataformas, realização do teste de qualificação, discussões em fóruns, interação com os demais participantes e, finalmente, submissão de um teste de software. É importante ressaltar que as plataformas de CST analisadas oferecem uma compensação financeira para o participante vencedor da atividade de teste funcional.

Participaram destas atividades 15 estudantes finalistas de um curso de graduação em Ciência da Computação, sem experiência em atividades de CST. A coleta durou sete semanas durante o segundo semestre de 2018 e os estudantes realizaram as possíveis ações de um trabalhador na multidão em sua primeira contribuição em CST. Ainda, como fonte de dados, foi adotada a abordagem perceptiva e cognitiva, cujo objetivo é investigar as relações entre as características do ambiente e o comportamento dos indivíduos.

Conforme Bardin (BARDIN, 1977) a análise de conteúdo está organizada em três fases: 1) pré-análise, 2) exploração do material e 3) tratamento dos resultados, inferência e interpretação. Na etapa de pré-análise foram sistematizadas as ideias iniciais e estabelecidos os indicadores para interpretação dos dados coletados. Os dados armazenados nos diários individuais foram analisados e organizados. Na segunda fase, a exploração do material consistiu na construção dos códigos, considerando unidades de registro nos textos analisados, e a classificação das informações em categorias simbólicas. Nesta etapa, foram selecionados trechos, dos diários, que expressam os códigos encontrados. A terceira etapa, será realizada dentro do contexto do trabalho e será responsável pelo tratamento dos resultados, inferência e interpretação, captando os conteúdos contidos em todo material coletado.

4. Resultados Preliminares

Os códigos gerados a partir da análise de conteúdo foram: Idioma, Comunicação, Homologação, Dispositivos, Tempo, Duplicidade, Pontuação, Experiência, Qualificação, Convite, Colaboração. As categorias apresentadas a seguir resultaram do processo de codificação dos diários. São apresentados alguns exemplos de trechos dos diários que serviram de base para a geração da categorização. Vale lembrar que não existem regras quanto à nomeação das categorias, nem para a quantidade.

A primeira categoria apresentada é ***Comunicação com a plataforma*** e resulta de trechos, como por exemplo: “*Como forma de comunicação com a plataforma, é possível entrar em contato através de e-mail ou chat, sempre no idioma inglês. Foi enviado e-mail questionando como era o processo para receber convites de projetos remunerados*”. Essa categoria abrange os dois primeiros códigos: “1” Idioma e “2” Comunicação.

⁶ <https://www.qsrinternational.com/>

Já a segunda categoria, **Homologação dos testes**, abrange os códigos “3” Homologação e “4” Dispositivos, e descreve a seleção dos seguintes trechos: “*Ocorre instabilidade em alguns sites de homologação, geralmente ocasionando bloqueio nos testes. Esses bloqueios são relatados como defeito e é necessário aguardar um feedback por parte do cliente para retomar os testes*”; “*Alguns clientes solicitam testes apenas para dispositivos móveis, com versões e navegadores bem específicos, o que dificulta na participação dos testes*”.

A terceira categoria, **Ciclos de testes**, é resultado de trechos analisados como por exemplo: “*Geralmente os projetos expiram em poucas horas, tendo seu ciclo de testes de no máximo dois dias*”, expondo o código “5” Tempo: “*Nos projetos remunerados há maior concorrência entre os participantes, então ocorrem relatos de defeitos duplicados. Quanto mais tempo o participante leva para iniciar seus testes, mais difícil será o relato de um defeito não duplicado*” referindo-se ao código “6” Duplicidade.

Ranking foi a quarta categoria obtida, por meio da seleção de trechos como por exemplo: “*A pontuação é iniciada com os defeitos do teste de qualificação, e a partir daí, são incluídos todos os defeitos relatados dentro dos projetos remunerados. Isso faz com que a pontuação aumente ou diminua de uma forma geral, e não por projeto*”. Outro trecho referente a categoria ranking: “*O ranking de pontuação leva em consideração os defeitos aprovados. Isso é facilmente alterado quando são reportados muitos defeitos inválidos, pois a pontuação diminui, conseqüentemente a classificação*”. Essa categoria abrange o código “7” Pontuação.

Já a quinta categoria apresentada é **Conhecimento**, no qual abrange o código “8” Experiência. Ela resulta de trechos, como por exemplo o selecionado “*Através das ferramentas de comunicação e interação com os demais participantes, pode ser possível absorver informações que servirão como auxílio nos testes dos projetos remunerados*”.

A sexta categoria, **Avaliação do novo participante**, refere-se aos códigos “9” Qualificação e “10” Convite. É resultado dos trechos “*É necessário realizar um teste de capacitação, inclusive com o intuito de proporcionar ao participante uma breve experiência do funcionamento da plataforma e no relato de defeitos*” e “*Os convites para projetos remunerados podem ocorrer de acordo com as informações prestadas no cadastro e pela pontuação obtida no teste de qualificação*”.

Concluindo, a lista das categorias finais, a **Participação na plataforma** resulta do trecho “*Contribuir na comunidade de teste, gerando conteúdos em forma de artigos e fóruns, demonstra o interesse do participante na participação de novos projetos remunerados*”. Essa categoria relaciona-se ao código “11” Colaboração.

5. Discussão Preliminar

Motivação e personalidade, são exemplos de atributos que impulsionam um participante a ingressar e contribuir em uma comunidade on-line (STOL e FITZGERALD, 2014). Também são considerados fatores como identificação com a comunidade, satisfação, envolvimento e confiança, que demonstram as intenções e comportamentos do participante (KITUR, CHI e SUH, 2008). A identificação e o envolvimento são demonstrados por meio de uma participação ativa nas atividades da comunidade de cada plataforma (STRAUB, GIMPEL, *et al.*, 2015).

Neste contexto, uma das barreiras identificadas relaciona-se ao posicionamento do participante no ranking geral de cada plataforma. Esse ranking denota experiência, participações em testes remunerados, conexões com os demais participantes, o quanto o

participante está contribuindo de forma positiva na comunidade. Essa barreira também se conecta com a participação ativa do usuário, pois gerar conteúdos de interesse sobre testes em CST e suas ferramentas é um exemplo de contribuição que agrega valor ao participante, tanto diante da plataforma quanto dos demais participantes da comunidade. Mesmo com a diversidade de participação em uma comunidade, é crucial que os participantes forneçam conteúdo suficiente e mantenham uma interação com os demais participantes para preservação da comunidade (ROHRMEIER, 2012).

Porém, a participação na plataforma também se demonstrou como uma barreira para os trabalhadores na multidão em CST. Estar conectado à plataforma não certifica que o usuário conseguirá convites para projetos remunerados, e se faz necessária uma participação ativa de forma diária. Identificar-se com as ferramentas disponíveis na plataforma, assim como com os demais participantes, colabora para que haja uma participação constante. Essa identificação social está relacionada com o tipo de consciência de cada participante. É definido como um sentimento de conexão intrínseca entre os membros e um senso de diferença coletivo de outros que não estão na mesma comunidade.

Outra barreira identificada foi quanto ao “grau” de conhecimento que o participante consegue obter ao fazer uso das ferramentas dentro de cada comunidade de CST. Nas comunidades sociais online, o valor é inerente à conexão entre os pares, a comunicação entre eles e a troca e combinação de conhecimento, informação e bens digitais (ROHRMEIER, 2012). A participação do usuário em ferramentas de comunicação como fóruns e chats, deve ocorrer sempre de forma a discutir assuntos que agreguem valor às experiências com as tarefas que vem sendo realizadas, neste caso, tarefas de teste de software. Comunicar-se com a plataforma, utilizando canais síncronos como o chat ou assíncronos como o e-mail, com o intuito de solucionar dúvidas quanto o escopo ou processo de teste, também gera um agregado de conhecimentos e ações objetivas para dúvidas futuras. Essa comunicação ocorre diretamente com os gestores de cada plataforma, sem envolver os demais participantes.

Além dessas ferramentas de comunicação, interagir com a comunidade da plataforma, pode resultar em troca de experiências e crescimento coletivo. Para Rohrmeier (ROHRMEIER, 2012), estar em contato com os demais participantes, em uma comunidade online, pode levar a mais interações, pois isso significa uma diversidade de fontes de informações disponíveis para o usuário focal. Ocorre, então, uma identificação social, que expressará um tipo de força de empatia para todos os contatos do participante. Ou seja, a identificação com outros participantes potencializa relacionamentos estáveis.

A realização da avaliação do novo participante por parte da plataforma foi outra barreira identificada. Essa etapa do processo consiste em comprovar as qualificações informadas no cadastro do participante na plataforma. Também está interligada ao posicionamento no ranking geral, pois cada defeito relatado no teste de entrada, conta pontos, negativos ou positivos. Este teste, serve como um primeiro contato do trabalhador na multidão com o processo de validação dos testes que cada plataforma exerce. A partir desta avaliação inicial, o participante já é inserido no ranking geral da plataforma, porém, não garante ao participante uma seleção para um convite de projeto de teste remunerado. Ao receber um convite para projeto remunerado, o participante deve comprometer-se com as tarefas propostas, buscando sempre entender seus escopos, pois as pontuações no ranking geralmente são acumulativas, e defeitos invalidados prejudicam o destaque positivo do participante.

6. Considerações finais

O *Crowdsourced Software Testing (CST)* auxilia no processo de validação da qualidade de um software por meio de testes realizados pela multidão de trabalhadores online. Porém, os trabalhadores deste modelo enfrentam algumas barreiras para executarem tais atividades. Foi possível identificar, preliminarmente, algumas barreiras relacionadas a comunicação e participação na plataforma. Para um trabalhador na multidão em atividades de CST receber um convite para atividades remuneradas é fundamental para manter-se ativo na plataforma, comunicar-se com seus pares e com a plataforma e isso, pode auxiliar e guiar o trabalhador durante sua participação nos projetos de teste. Experiências e conhecimento coletivo agregam valor na qualidade dos testes realizados.

Considerando o estágio exploratório e preliminar do trabalho, os resultados não podem ser generalizados pois as atividades foram realizadas com apenas 15 participantes inexperientes em atividades de CST, e finalização da análise dos dados. Como trabalho futuro realizar-se-á a terceira etapa da Análise de Conteúdo e refinar a análise de dados para, se possível, explorar como ocorre a comunicação/interação com a plataforma. A partir destas categorias finais se faz necessário elaborar um conjunto de recomendações para que o trabalhador da multidão possa superar as dificuldades encontradas em atividades CST.

Referências

- BARDIN, L. Analysis of content. Lisbon: [s.n.], 1977.
- HILL, W. C. et al. Edit wear and read wear. Conference on Human Factors in Computing Systems. New York: [s.n.]. 1992. p. 3–9.
- KITTUR, A.; CHI, E. H.; SUH, B. Crowdsourcing user studies with Mechanical Turk. Human factors in computing systems. [S.l.]: [s.n.]. 2008. p. 453–456.
- LU, X.; PHANG, W.; YU, J. Encouraging participation in virtual communities through usability and sociability development: An empirical investigation. Special Interest Group on Management Information Systems, 42, n. 3, Ago 2011. 96-114.
- MAO, K. et al. A Survey of the Use of Crowdsourcing in Software Engineering. Journal of Systems and Software, 126, n. 1, Abr 2016. 57-84.
- ROHRMEIER, P. Social networks and online communities—managing user acquisition, activation and retention. Technischen Universität München. München. 2012. (322p.).
- SAREMI, R. L.; YANG, Y. Dynamic simulation of software workers and task completion. 2nd International Workshop on CrowdSourcing in Software Engineering. Florence: [s.n.]. Mai 2015. p. 17-23.
- STOL, K.; FITZGERALD, B. Researching crowdsourcing software development: perspectives and concerns. 1st International Workshop on CrowdSourcing in Software Engineering. Hyderabad: [s.n.]. 2014. p. 7-10.
- STRAUB, T. et al. How (not) to Incent Crowd Workers - Payment Schemes and Feedback in Crowdsourcing. Business & Information Systems Engineering, 57, n. 3, Jun 2015. 167-179.
- WU, W.; LI, W.; TSA, W. T. An evaluation framework for software crowdsourcing. Frontiers of Computer Science, v. 7, n. 5, p. 694-709, Out 2013.
- YANG, Y. et al. Who Should Take This Task?: Dynamic Decision Support for Crowd Workers. 10th International Symposium on Empirical Software Engineering and Measurement. [S.l.]: [s.n.]. 2016. p. 8.