

COVID-19 na Indústria de Software: uma Revisão de Literatura Cinza auxiliada por uma Ferramenta de Web-Scraping

Wiliane Maria A. S. de Souza¹, Wylliams B. Santos¹

¹Universidade de Pernambuco (UPE) – Caruaru, PE – Brasil

{wiliane.maria,wbs}@upe.br

Abstract. *The SARS-CoV-2 pandemic brings challenges for the software industry. We believe that, in this context, companies are implementing practices and tools to overcome those. we performed a Gray Literature Review for which a search tool was developed in the agile software development context with web scraping techniques. We use the Design Science method and bring results including 63 practices and tools, with the communication ones standing out; challenges such as communication and collaboration difficulties, disruption, decreased productivity, infrastructure, distance, personal conflicts and health and well-being issues.*

Resumo. *A pandemia do SARS-CoV-2 traz desafios para a indústria de software. Acreditamos que, nesse contexto, as empresas implementam práticas e ferramentas para contorná-los. Objetivando identificá-los, realizamos uma Revisão de Literatura Cinza para a qual foi desenvolvida uma ferramenta de busca no contexto ágil de software com técnicas de web scraping. Utilizamos o método de Design Science e trazemos resultados que incluem 63 práticas e ferramentas, destacando-se as de comunicação; desafios como dificuldades de comunicação e colaboração, interrupção, diminuição de produtividade, infraestrutura, distância, conflitos pessoais e problemas de saúde e bem estar.*

1. Introdução

A pandemia do SARS-CoV-2, vírus causador da COVID-19, anunciada em 2020 pela Organização Mundial de Saúde [WHO 2020], trouxe barreiras para a indústria. Acreditamos que organizações na Indústria de *Software* (IS) vêm, em seu contexto, tomando atitudes para contornar os desafios advindos das medidas restritivas recomendadas, incluindo o isolamento social. É interessante entender quais os problemas se fazem presentes e as boas práticas que vêm sendo empregadas por elas, por meio da Literatura Cinza (LC).

Para a busca de LC ainda não existe um processo tal qual o de busca de literatura branca, que abrange plataformas que suportam *strings* de busca, filtros e exportação dos resultados de maneira unificada, realizando tal busca de maneira automática e eficiente. Nesse trabalho, com o uso de técnicas de raspagem de dados (*Web Scraping*), desenvolvemos tal ferramenta, que será utilizada em uma revisão de LC, com o foco em compreender os desafios e boas práticas da indústria de software durante a pandemia.

2. Fundamentação Teórica

A LC é definida como múltiplos tipos de documentos produzidos em todos os níveis governamentais, acadêmicos e empresariais em formato impresso ou eletrônico, desde que não sejam controlados por editoras comerciais [Schöpfel 2010]. Sua utilização em trabalhos no campo de ES também está associada ao interesse em incorporar o ponto de vista de indivíduos do campo prático [Kamei et al. 2021] [Neto et al. 2020].

A técnica de *Web Scraping* é utilizada para extrair dados de múltiplos *websites* para um único banco de dados ou planilha, facilitando a análise e visualização dos dados. Ela envolve a captura e extração de informações de uma página *web*, onde, para que os dados possam ser extraídos, primeiro é capturado todo o conteúdo da página e, posteriormente, extraídas as informações desejadas [Fiallos et al. 2020]. Estudos como [Georgiou et al. 2020] e [Fiallos e Jimenes 2019] realizam *Web Scraping* com ferramentas como o *Selenium* e *BeautifulSoup* na linguagem *Python*.

3. Metodologia e Solução Proposta

Foi utilizado o método de pesquisa *Design Science* (DS), que afirma que ao produzir coisas úteis estamos produzindo conhecimento científico e enfatiza a conexão entre o conhecimento científico e o campo prático [Wieringa 2009]. A primeira etapa consiste na Investigação do Problema, abordada na introdução deste trabalho. A seguir, apresentamos as implementações das etapas seguintes implementadas nesse trabalho.

Foi desenvolvida uma ferramenta de *Web Scraping*, detalhada na Figura 1, com finalidade de coleta de dados da LC no contexto de desenvolvimento ágil de *software*. A revisão de LC foi guiada pelas seguintes perguntas de pesquisa: **RQ1:** Quais os desafios enfrentados pelas empresas de desenvolvimento de *software* durante a pandemia? **RQ2:** Quais as práticas e ferramentas implementadas pelas empresas de desenvolvimento de *software* para contornar os desafios enfrentados durante a pandemia?

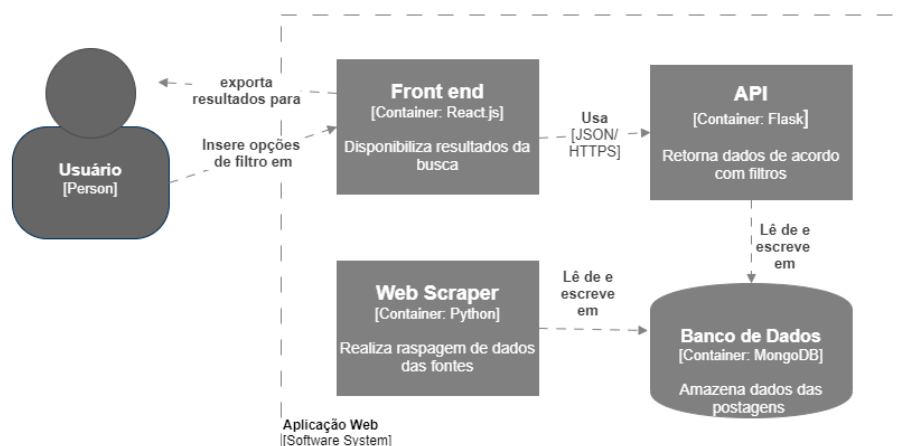


Figura 1. Diagrama de container C4 da Aplicação Web.

A ferramenta de *Web Scraping* está disponível publicamente por meio do endereço eletrônico: <https://search-for-grey.netlify.app>. Ela utiliza como fontes *websites* dentro do contexto de desenvolvimento ágil de *software*: Scrum.org, AgileConnection.com, AgileAlliance.org e o ScrumAlliance.org. Contudo, esta última foi incluída em uma atualização posterior da ferramenta e, portanto, não foi considerada na

revisão aqui apresentada. Tais fontes são *websites* que emitem certificações reconhecidas no contexto ágil de *software*.

A seleção das publicações retornadas pela ferramenta foi baseada no critério de inclusão (CI): Literatura cinza que relaciona gestão de projetos de *software* no contexto do COVID-19. Foram excluídas publicações a partir de três critérios de exclusão (CE). (CE1): Não é uma publicação que relaciona gestão de projetos de *software* e COVID-19; (CE2): Não é uma publicação relacionada a gestão de projetos de *software*; (CE3): Não é uma publicação escrita em inglês.

A partir da *string* de busca¹, a ferramenta retornou 56 resultados compatíveis. Os resultados foram individualmente analisados por dois dos autores, que aplicaram os critérios de inclusão e exclusão. O período utilizado para a busca foi entre janeiro de 2020 e março de 2021. Para a análise e síntese dos dados dos 19 resultados incluídos, foi utilizada uma adaptação do guia para condução de sínteses narrativas em revisões sistemáticas proposto por [Popay et al. 2006], sendo composto pelas etapas de desenvolvimento da síntese preliminar e exploração dos relacionamentos entre os dados. Na primeira etapa foram identificados os desafios, ferramentas e práticas presentes em cada um deles. Ao fim dessa etapa foi gerada a síntese preliminar dos achados da análise, resumindo os desafios encontrados e também as ferramentas e práticas implementadas.

4. Resultados da revisão

4.1. RQ1

A partir das publicações analisadas foram identificados os seguintes desafios: dificuldades de comunicação e colaboração; interrupção; diminuição de produtividade; falta de infraestrutura; distância; conflitos pessoais; problemas de saúde e bem estar.

4.2. RQ2

Foi identificado um conjunto de 63 práticas e ferramentas, das quais cerca de 35% representam artefatos, frameworks e práticas ágeis, refletindo o contexto de desenvolvimento ágil de *software*. Se destacam o framework Scrum e o artefato ágil de *sprints*. Em relação às práticas, as mais mencionadas foram *retrospectives*, *daily Scrums* e *Scrum of scrums meeting*.

As práticas e ferramentas de comunicação e colaboração representam pouco mais de 25% do total, destacando-se as práticas de comunicação, que compreendem 50% desse número. As ferramentas de colaboração mencionadas foram descritas como capazes de auxiliar no planejamento e realização de *retrospectives* em conjunto, mantendo transparência no trabalho. Destacam-se também a aparição de habilidades interpessoais (*soft skills* e *leadership skills*), que somam cerca de 13% do total, sendo destacadas como importantes no contexto da pandemia. Entre as atitudes de time citadas, aparecem o “awareness agreement”, definido como um acordo interno ao time em relação ao que precisa ser feito e, semelhante a ele, o “team ‘ways of working’”, que se refere a criação de um acordo criado pelo time em relação a como o time irá funcionar.

¹ [(“software development” OR “software Project” OR “software engineering” OR “software team” OR “project management” OR “project manager” OR “Agile” OR “Extreme Programming” OR “XP” OR “Lean Software Development” OR “SCRUM” OR “Kanban”) AND (“SARS-CoV-2” OR “COVID-19” OR “Coronavirus” OR “2019-ncov”)]

5. Conclusões e considerações finais

Foi conduzida uma revisão de LC com a finalidade de identificar desafios enfrentados pelas empresas de desenvolvimento de *software* durante a pandemia e quais práticas e ferramentas utilizadas por elas para contorná-los. Para isso, foi desenvolvida e utilizada uma ferramenta de *web scraping*. Foram identificados desafios como: dificuldades de comunicação e colaboração; disrupção; diminuição de produtividade; falta de infraestrutura; distância; conflitos pessoais; problemas de saúde e bem estar. Identificamos também um conjunto de 63 práticas e ferramentas, de maneira coerente com os desafios relatados nas publicações.

Concluimos, portanto, que as práticas e ferramentas destacadas são as de colaboração e comunicação, possivelmente refletindo os desafios ligados à disrupção, diminuição de produtividade e distância. Entende-se que esse estudo possui suas limitações tais quais o escopo limitado da ferramenta ao contexto ágil e o baixo número de resultados retornados até então.

Referências

- Asikri, M., Chaib, H., Salah-ddine, K. (2020). "Using Web Scraping In A Knowledge Environment To Build Ontologies Using Python And Scrapy". In: European Journal of Translational and Clinical Medicine.
- Fiallos, A. and Jimenes, K. (2019) "Using Reddit Data for Multi-Label Text Classification of Twitter Users Interests", In: 2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG).
- Georgiou, K., Mittas, N., Angelis, L. and Chatzigeorgiou, A. (2020) "A Study of Knowledge Sharing related to Covid-19 Pandemic in Stack Overflow". arXiv preprint arXiv:2004.09495.
- Kamei, F. K., Wiese, I., Lima, C., Polato, I., Nepomuceno, V., Ferreira, W., Ribeiro, M., Pena, C., Cartaxo, B., Pinto, G. and Soares, S. (2021). "Grey Literature in Software Engineering: A Critical Review", In: Information and Software Technology.
- Neto, G. T. G., Santos, W. B., Endo P. T. and Fagundes, R. A. A. (2019) "Multivocal literature reviews in software engineering: Preliminary findings from a tertiary study," ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2019, pp. 1-6, doi: 10.1109/ESEM.2019.8870142.
- Organization, World Health. (2020) "WHO announces COVID-19 outbreak a pandemic", <http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>. Acesso em: 02 abr. 2020.
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., and Duffy, S. (2006) "Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A product from the ESRC Methods Programme". 10.13140/2.1.1018.4643.
- Wieringa, R. (2009), "Design Science as Nested Problem Solving". In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09.