

Avaliando um SI Colaborativo para Apoiar Revisões e Mapeamentos Sistemáticos da Literatura

Evaluating a Collaborative IS to Support Systematic Literature Reviews and Systematic Mapping Studies

**Thiago Prado de Campos^{1,2}, Eduardo Filgueiras Damasceno¹,
Natasha Malveira Costa Valentim²**

¹Universidade Tecnológica Federal do Paraná (UTFPR)
Londrina/Cornélio Procópio, PR – Brasil

²DInf – Universidade Federal do Paraná (UFPR)
Curitiba, PR – Brasil

contato@thiagotpc.com, damasceno@utfpr.edu.br, natasha@inf.ufpr.br

Abstract. *This article presents a study to evaluate the quality of the Porifera collaborative tool, which supports Systematic Literature Review (SLR) and Systematic Mapping Study (SMS) processes. The study revealed a high perceived usefulness and ease of use of the tool and identified points for improving the interface, such as user feedback and error prevention. The quantitative and qualitative results suggest that Porifera is a valuable tool for learning and performing collaborative SLR and SMS. Additionally, the article compares Porifera with other related tools, including a discussion on how these other tools were evaluated.*

Keywords. *Systematic Literature Review; Systematic Mapping Study; Collaborative System; Software Evaluation; User Feedback*

Resumo. *Este artigo apresenta um estudo para avaliar a qualidade da ferramenta colaborativa Porifera, que apoia processos de Revisão Sistemática da Literatura (RSL) e Mapeamento Sistemático da Literatura (MSL). O estudo revelou alta utilidade percebida e facilidade de uso da ferramenta, além de identificar pontos para melhorar a interface, como o feedback ao usuário e prevenção de erros. Os resultados quantitativos e qualitativos sugerem que Porifera é uma ferramenta valiosa para a aprendizagem e realização de RSL e MSL de forma colaborativa. Além disso, o artigo faz uma comparação entre Porifera e outras ferramentas relacionadas, incluindo uma discussão sobre a forma que estas outras ferramentas foram avaliadas.*

Palavras-Chave. *Revisão Sistemática da Literatura; Mapeamento Sistemático da Literatura; Sistema Colaborativo; Avaliação de Software; Feedback do Usuário*

1. Introdução

Uma Revisão Sistemática da Literatura (RSL) é um meio de identificar, avaliar e interpretar todas as pesquisas disponíveis relevantes para uma questão de pesquisa [Kitchenham and Charters 2007]. Um Mapeamento Sistemático da Literatura (MSL) é usado para construir um esquema de classificação e estruturar um campo de interesse [Petersen et al. 2008]. Ambos são estudos secundários que se debruçam sobre a literatura existente com objetivo de identificar lacunas de pesquisa [Petersen et al. 2008, Kitchenham et al. 2010]. Cada área de pesquisa possui um guia particular para execução de RSL [Okoli and Schabram 2010]. Em geral, o processo costuma ser composto por Planejamento (definição das questões de pesquisa e do protocolo), Condução (identificação da pesquisa, a seleção dos estudos, avaliação da qualidade dos estudos, extração e síntese de dados) e Disseminação (escrita de relatórios e divulgação científica) [Kitchenham and Charters 2007]. Seus procedimentos bem definidos permitem uma execução objetiva e tornam o processo auditável e repetível de forma que outros pesquisadores, ao seguirem o mesmo protocolo, possam selecionar os mesmos estudos primários e chegar às mesmas conclusões ou classificação [Kitchenham et al. 2015]. Como desvantagem, estes estudos secundários exigem um esforço consideravelmente maior do que outras revisões não sistemáticas [Kitchenham and Charters 2007, Kitchenham et al. 2010]. MSL e RSL são processos complementares [Kitchenham and Charters 2007], portanto, neste artigo, para facilitar a fluidez da leitura, usaremos apenas o termo “RSL”, deixando implícito a referência também à MSL.

Um suporte computacional pode facilitar o trabalho e possibilitar maior qualidade no processo de execução de RSL [Hernandes et al. 2012]. Um software pode fornecer apoio valioso em diversos aspectos, como aqueles que envolvem o armazenamento, gerenciamento, validação e análise do que pode ser uma grande quantidade de dados [Kitchenham et al. 2015]. Parte destas atividades pode ser propensas a erros e demoradas quando realizadas manualmente [Kitchenham et al. 2015].

De um levantamento realizado com 131 participantes - pesquisadores da área de Engenharia de Software (ES) - apenas 53 usavam uma ferramenta para apoiar o processo de pesquisa, sendo que a maioria usava como apoio planilhas eletrônicas (22 pessoas) ou gerenciadores de referências bibliográficas (28 pessoas) [Al-Zubidy and Carver 2019]. Além disso, vale ressaltar que tais softwares não atendem especificamente ao fluxo de atividades de uma RSL, trazendo algumas desvantagens, como: a) o enviesamento da avaliação dos pesquisadores por visualizarem a avaliação feita por colegas no mesmo documento compartilhado ou; b) a possibilidade de exclusão ou modificação não intencional de qualquer informação no documento, do próprio documento ou registro da coleção de publicações.

Com objetivo de superar estas e outras desvantagens, ferramentas específicas para RSL foram criadas, abordando diferentes etapas das revisões ou mapeamentos sistemáticos, tais como elaboração do protocolo de pesquisa, automação da busca, seleção de estudos, avaliação da qualidade, extração de dados, análise automatizada, análise de texto e metadados, gestão de documentos, escrita de relatório e colaboração. No levantamento de Al-Zubidy and Carver [2019], poucas ferramentas específicas são usadas -

apenas nove pessoas indicaram usar ferramentas específicas e apenas quatro delas foram citadas. O estudo indica que ainda é necessário um apoio ferramental específico e sugere que o fato de poucas serem usadas talvez se dê em razão de uma série de limitações observadas nas ferramentas existentes, como, por exemplo, a de que ainda é necessário realizar muito trabalho manual, como para remoção de falso positivos e limpeza de referências.

O site Systematic Review Toolbox (SR Toolbox)¹ [Marshall and Brereton 2015] possui coleção de mais de 200 ferramentas de apoio para RSL. Em meio a tantas soluções, algumas são específicas para atender o domínio da Engenharia de Software (ES) e áreas correlatas. Usar uma ferramenta específica para apoiar RSL pode trazer benefícios, como: redução do esforço necessário; agilidade na execução do processo; torna o processo mais claro, preciso, organizado e validável; facilita a colaboração e; melhora a apresentação dos resultados [Marshall 2016].

Dentre os requisitos desejados em ferramentas de apoio a RSL [Marshall et al. 2014, Hassler et al. 2016, Al-Zubidy et al. 2017], o aspecto da colaboração é um dos mais solicitados pelos autores deste tipo de pesquisa. Em um *workshop* com autores experientes e novatos [Hassler et al. 2016], este aspecto foi o mais votado como desejado pelos participantes, sendo requisitado em diversas etapas do processo, principalmente na definição do protocolo e na realização de pesquisa e seleção de estudos. Além do requisito de colaboração, a maioria dos recursos de alta prioridade não são bem suportados pelas ferramentas atuais [Hassler et al. 2016]. De forma geral, autores novatos dão mais importância a recursos que auxiliam na execução de tarefas operacionais primárias (como análise, rastreamento, estimativa de tempo, colaboração e extração de dados) e, os mais experientes, em recursos que auxiliam em atividades táticas (como colaboração, desenvolvimento de protocolo e validação) [Hassler et al. 2016].

O processo de RSL é colaborativo por sua própria natureza [Al-Zubidy et al. 2017]. Para reduzir os efeitos do viés de um único pesquisador, vários pesquisadores devem executar colaborativamente o processo, revisando e sintetizando pesquisas publicadas na literatura [Al-Zubidy et al. 2017]. A colaboração é necessária em todas as fases, desde o desenvolvimento de protocolo, passando por pesquisa, seleção, resolução de conflitos, extrações de dados, síntese e concluindo na redação de relatórios [Al-Zubidy et al. 2017]. Portanto, há necessidade de facilitar a colaboração entre membros de equipe próximos ou distribuídos [Al-Zubidy et al. 2017].

Um exemplo do aspecto da colaboração na etapa de seleção de estudos primários de uma RSL está presente nos critérios de avaliação da *feature* “F3-F04” do conjunto proposto por Marshall et al. [2014]. Espera-se de uma ferramenta ofereça suporte para um processo de seleção em várias etapas (por exemplo, título/resumo e, em seguida, artigo completo), por meio da aplicação individual de critérios de inclusão ou exclusão da parte de cada pesquisador. Havendo desacordo, o sistema deve permitir a discussão e fornecer uma facilidade para reconciliar discordâncias, como sinalizar uma decisão em consenso ou deixar claro outra regra usada para resolver o desacordo - ver [Petersen and Ali 2011]. Importante que isto ocorra sem perder o histórico da avaliação de cada pesquisador, garantindo a transparência e rastreabilidade da decisão final.

¹<http://systematicreviewtools.com/>

Uma vez que não foi encontrada uma ferramenta colaborativa com as características citadas acima disponível para uso, Porifera foi proposta como solução. Porifera é um Sistema de Informação (SI) para gestão de dados, informação e conhecimento. Ele organiza registros de publicações permitindo que sejam interpretados e processados, gerando informação e conhecimento, advindo da integração entre as revisões e colaboração dos pesquisadores, possibilitando tomada de decisão. Atua mais especificamente nas etapas de Planejamento e Condução com foco na definição do protocolo e, principalmente, na seleção de estudos. Porifera é voltado para um mundo globalizado e conectado, permitindo a realização de uma pesquisa assíncrona, com mobilidade, colaboração, compartilhamento de conhecimento, flexibilidade, dentre outros aspectos. Permite em certo grau a interoperabilidade com outros sistemas e a integração entre pessoas e tecnologia, facilitando a aplicação de métodos e práticas de RSL por pesquisadores de diversos níveis de formação.

Um apoio inadequado provoca aumento do esforço necessário para realização de uma RSL e adiciona barreiras a novos pesquisadores que desejam realizar este tipo de revisão [Hassler et al. 2016]. Por esta razão, é necessário garantir que as ferramentas propostas para apoiar uma RSL cubram parte das funcionalidades mais desejadas para este tipo de ferramenta e que sua implementação tenha boa usabilidade e aceitação para seu público-alvo. Portanto, um estudo foi realizado com objetivo de responder a seguinte questão de pesquisa: “Porifera é uma ferramenta colaborativa útil, eficaz e fácil de usar?”. Por isso, Porifera foi submetida para avaliação de sua qualidade por 11 estudantes da Universidade Federal do Paraná (UFPR) que participaram de uma disciplina com foco em pesquisa experimental nas áreas de ES e Interação Humano-Computador (IHC). Eles utilizaram Porifera por duas semanas durante um trabalho desta disciplina, onde tinham que definir o protocolo de uma RSL e realizar o processo parcial da seleção de estudos primários. Após isso, os alunos responderam um questionário contendo sentenças do *Technology Acceptance Model* (TAM) [Davis 1993] e questões abertas com objetivo de identificar suas percepções e melhorias de Porifera. Este artigo, que é uma extensão de Campos et al. [2022], apresenta de forma quantitativa e qualitativa resultados sobre utilidade percebida, facilidade de uso, intenção de uso e eficácia de Porifera. Também traz problemas e pontos de melhoria identificados durante a avaliação.

A seção a seguir apresenta trabalhos relacionados. A Seção 3, apresenta Porifera de forma breve e alguns dos seus principais recursos. A Seção 4 detalha a avaliação da qualidade realizada e a Seção 5 apresenta os resultados obtidos. Na Seção 6 são apresentadas conclusões e informações sobre trabalhos futuros.

2. TRABALHOS RELACIONADOS

Esta seção é dividida em três partes: primeiro apresenta as principais ferramentas encontradas com características parecidas a Porifera. Depois apresenta os métodos que cada uma delas usou para avaliar sua qualidade. Em seguida, cada um dos métodos de avaliação é explicado e um quadro relacionando as ferramentas e os métodos usados é apresentado.

2.1. Ferramentas de Apoio para RSL Colaborativos

Dada a necessidade de realização de um MSL colaborativo, principalmente na etapa de seleção de estudos, algumas ferramentas de apoio foram pesquisadas e testadas durante o segundo semestre do ano de 2020. Primeiramente, foram analisadas as ferramentas que apareceram nos levantamentos de requisitos de Marshall et al. [2014] e Al-Zubidy et al. [2019]. Depois, foi realizada uma busca no catálogo SR Toolbox, com objetivo de identificar ferramentas colaborativas de apoio para RSL no processo de seleção de estudos (filtro ativado em recursos de “seleção de estudos” e “colaboração”). Recentemente, o catálogo SR Toolbox foi atualizado e sua interface de pesquisa modificou os filtros, refletindo nova categorização proposta por Sutton et al. [2019]. Das ferramentas encontradas, foram testadas apenas ferramentas gratuitas, com foco na disciplina de ES ou multidisciplinar, e que fornecessem uma URL para uso (sem a necessidade de configurar um ambiente de execução).

Marshall et al. [2014] analisou as ferramentas SLuRp [Bowes et al. 2012], SLRTOOL [Barn et al. 2014], StArt [Hernandes et al. 2012] e SLR-Tool [Fernández-Sáez et al. 2010]. Al-Zubidy et al. [2019] analisou, além dessas, SESRA [Molléri and Benitti 2015] e Parsifal [Parsifal Ltd. 2018]. Por meio da pesquisa no catálogo SR Toolbox foram identificadas: CADIMA [Kohl et al. 2018], Parsifal [Parsifal Ltd. 2018], Rayyan [Ouzzani et al. 2016], SESRA [Molléri and Benitti 2015] e Thoth [Marchezan et al. 2019]. Uma última ferramenta foi encontrada enquanto se pesquisava por artigos acadêmicos sobre as ferramentas listadas em SR Toolbox. Trata-se de Buhos [Bustos Navarrete et al. 2018].

Algumas ferramentas requerem configuração e instalação em ambiente específico para sua utilização. É o caso de SLRTOOL, StArt, SLR-Tool e Buhos. Por esta razão, tais ferramentas não foram testadas.

SLuRp [Bowes et al. 2012] não foi testada, pois, embora tenha indicado uma URL, não estava acessível. De acordo com Marshall et al. [2014], SLuRp identifica discordâncias entre avaliações dos pesquisadores e, quando isto ocorre, para resolver a disputa é necessário a moderação por um usuário de fora do conflito, que atua como mediador.

ReLiS [Bigendako and Syriani 2018] também não foi testada, pois nela as definições da RSL são feitas por meio de um arquivo de linguagem específica de domínio que deve ser carregado no sistema, e não por uma interface gráfica. Este fato, somado ao pouco entendimento sobre o modelo, impediu testar a ferramenta. ReLiS foi avaliada recentemente por Stefanovic et al. [2021] e segundo os autores, ela atende os critérios de avaliação da *feature* F3-F04 de Marshall et al. [2014].

As ferramentas testadas foram: CADIMA [Kohl et al. 2018], Parsifal [Parsifal Ltd. 2018], Rayyan [Ouzzani et al. 2016], SESRA [Molléri and Benitti 2015] e Thoth [Marchezan et al. 2019]. O objetivo do teste foi identificar duas características-chaves destas ferramentas: 1) seleção de estudos em dois filtros por meio de aplicação de critério de inclusão ou exclusão individualmente por cada pesquisador e; 2) sinalização e tratamento de desacordos, possibilitando aos pesquisadores tomarem decisões em consenso quando não houver concordância sobre os critérios aplicados individualmente.

Na ferramenta CADIMA [Kohl et al. 2018], para cada publicação é necessário indicar para todos os critérios definidos no planejamento, se o critério se aplica, não se aplica ou se não está claro para o avaliador. Isto torna o processo de seleção mais lento, ainda que seja possível marcar uma mesma decisão para todos os critérios.

Em Parsifal [Parsifal Ltd. 2018], quando um pesquisador avalia uma publicação utilizando um dos critérios de inclusão ou exclusão, esta avaliação passa a ser a decisão do grupo sobre a publicação; se outro pesquisador atualizar o critério aplicado sobre aquela mesma publicação, a última modificação é a que fica registrada. Não há espaço para cada pesquisador aplicar um critério sob seu ponto de vista, permitindo existir concordância ou desacordo na avaliação. Além disso, não permite mais de uma etapa de filtro, ou seja, a inclusão ou exclusão da publicação é realizada em uma única fase.

Rayyan [Ouzzani et al. 2016] apresenta um modo cego (*Blind ON*) e não cego (*Blind OFF*). No primeiro, não se vê a avaliação dos colegas e no segundo, as avaliações de todos são exibidas. A decisão sobre cada publicação é tomada por opções *Include/Exclude/Maybe* e indicar a razão é opcional. Ou seja, ela não permite a definição de critérios de inclusão ou de exclusão específicos. Registra-se apenas a natureza do critério (inclusão/exclusão).

SESRA [Molléri and Benitti 2015] é bem aderente às principais diretrizes para processos de RSL na ES, incluindo na etapa de planejamento verificar a necessidade de iniciar uma revisão e a avaliação do protocolo de revisão. Durante o planejamento, permite escolher entre três estratégias de seleção de estudos: dois ou mais pesquisadores avaliam cada publicação; um pesquisador primário faz uma seleção de forma completa e uma amostra é disponibilizada a um revisor para indicar a confiabilidade e; classificação automática com base na qualidade dos estudos. No primeiro caso, quando dois pesquisadores divergem no critério de avaliação, a ferramenta indica que um terceiro deve mediar a decisão. Porém, com dois pesquisadores apenas cadastrados, não parece haver modo de resolver o desacordo.

Thoth [Marchezan et al. 2019] é uma ferramenta de código aberto e uso gratuito que permite que cada pesquisador envolvido no projeto realize sua avaliação para cada publicação. Thoth possui recursos para desenvolvimento de um protocolo de pesquisa, seleção de estudos, avaliação da qualidade, extração de dados, síntese de dados e relatório [Marchezan et al. 2019]. Durante o processo de seleção de estudos, a ferramenta não apresentou quais foram as avaliações concordantes e discordantes, bem como que ação tomar quando há um desacordo. Também não há uma distinção de avaliação por filtros, cada pesquisador parece fazer sua avaliação apenas uma vez.

Buhos [Bustos Navarrete et al. 2018] é um projeto de software livre que deve ser instalado em ambiente adequado. Uma URL apenas de demonstração está disponível e por meio dela e do seu manual de usuário foi possível verificar que a seleção de estudos primários ocorre em múltiplas etapas. Cada pesquisador faz sua escolha entre incluir ou excluir o documento e é o administrador que seleciona quais seguirão para próxima etapa.

No geral, as ferramentas testadas têm seus méritos e vantagens para apoiar uma RSL, mas não foram encontradas nelas as características desejadas no processo de seleção

de estudos e forma colaborativa. Espera-se que a seleção de estudos possa ser realizada em duas fases, por meio da aplicação individual de critérios de inclusão ou exclusão da parte de cada pesquisador. Havendo desacordo, o sistema deve permitir a discussão e sinalizar uma decisão em consenso (ou deixar claro outra regra usada para resolver o desacordo), sem perder o histórico da avaliação de cada pesquisador, garantindo a transparência e rastreabilidade da decisão final.

2.2. Avaliações das Ferramentas de RSL

A maioria dos trabalhos que apresenta uma ferramenta de apoio para RSL apenas mostra um exemplo das funcionalidades ou indica que pequenos testes de uso estão sendo ou serão realizados [Marshall 2016]. SLR-Tool [Fernández-Sáez et al. 2010] [Hinderks et al. 2020], por exemplo, está sendo usada e testada em um grupo de pesquisa do Departamento de Linguagens e Sistemas de Computação da Universidade de Sevilha, mas uma validação ainda será realizada no futuro. SLuRp [Bowes et al. 2012], SLRTool [Barn et al. 2014], CADIMA [Kohl et al. 2018] e Parsifal [Parsifal Ltd. 2018] são ferramentas das quais não se encontrou publicações contendo avaliações de qualidade.

O trabalho que apresenta ReLiS [Bigendako and Syriani 2018] relata que foi realizada uma avaliação qualitativa informal preliminar, para determinar a implementação correta do sistema e utilidade de decisões em relação à arquitetura da ferramenta. Sem apresentar detalhes de como aconteceu a avaliação, os autores disseram que oito grupos de 3 a 4 alunos usaram ReLiS em disciplina relacionada a métodos empíricos em ES e que o *feedback* dos alunos e os registros do sistema foram usados para aprimorar a ferramenta.

StArt [Hernandes et al. 2012] foi avaliada por meio de um questionário baseado no TAM [Davis 1993]. O questionário foi planejado sob o paradigma *Goal-Question-Metric* (GQM) [Basili and Caldiera 2000]. Os participantes eram alunos de pós-graduação em Ciência da Computação da Universidade Federal de São Carlos (UFSCar), matriculados na disciplina de Metodologia da Pesquisa. A avaliação contou com 14 alunos em uma primeira avaliação e 35 em uma segunda avaliação. O uso do TAM tornou a avaliação rápida e objetiva [Hernandes et al. 2012]. Além das sentenças relacionadas ao TAM, o questionário continha campos abertos para comentários.

SESRA [Molléri and Benitti 2015] foi avaliada usando GQM sob critérios de produtividade e confiabilidade, misturando métricas objetivas e subjetivas. Já as métricas subjetivas foram investigadas por meio de uma pesquisa com profissionais e pesquisadores com experiência de uso na ferramenta, coletando respostas em questionários com escala Likert² e as métricas objetivas foram investigadas por meio de procedimentos experimentais conduzidos nos recursos automatizados de SESRA. Coletou-se *feedbacks* de usuários durante avaliação subjetiva.

Rayyan [Ouzzani et al. 2016] é uma ferramenta que possui um processo semi-automatizado para realizar a triagem inicial de resumos e títulos. Trata-se de um algoritmo que aprende com as decisões de inclusão e exclusão do usuário e tenta prever a avaliação do usuário sobre uma publicação, sugerindo uma classificação para os trabalhos ainda não

²Concordo Amplamente, Concordo Parcialmente, Sem Opinião, Discordo Parcialmente e Discordo Amplamente

verificados. Por isso, o processo de avaliação de Rayyan incluiu a avaliação da precisão do classificador, tomando como base duas RSLs realizadas e publicadas em Cochrane³. Um teste piloto foi realizado por dois desenvolvedores da ferramenta carregando duas revisões pré-existentes e, ao avaliar a precisão do algoritmo, puderam também avaliar de forma geral outros aspectos, como aparência do aplicativo e suas funcionalidades. Todos os comentários foram registrados por meio de uma função de *feedback* presente na ferramenta. Rayyan foi apresentada e disponibilizada para uso da comunidade em geral e, enquanto era usada, foi recebendo *feedbacks* dos usuários. A ferramenta inclui também um questionário de pesquisa que pode ser respondido a qualquer momento pelo usuário. Os autores relatam até a publicação do trabalho terem coletadas 66 respostas. Mas não foi exibida a composição deste questionário.

Thoth [Marchezan et al. 2019] faz uma avaliação comparativa entre si e Parsifal usando o *framework* proposto por Marshall et al. [2014]. Além disso, foi realizado também um quasi-experimento do qual participaram 12 pesquisadores e alunos de graduação e mestrado do curso de ES que já tivessem realizado anteriormente uma RSL. Os participantes foram divididos em grupos de duas pessoas e cada membro usou apenas uma ferramenta (Thoth ou Parsifal) por dois meses durante a realização de uma RSL. Depois, os participantes responderam um questionário com sete questões em escala Likert objetivando identificar se a ferramenta usada oferecia suporte satisfatório para realização de uma RSL e se o uso da ferramenta reduzia o esforço e complexidade da condução da RSL. Além disso, foram coletadas opiniões sobre vantagens e desvantagens das ferramentas.

Buhos [Bustos Navarrete et al. 2018] foi avaliado considerando três critérios: usabilidade, satisfação quanto a requisitos para RSL conforme metodologia proposta por Al-Zubidy et al. [2019] e; por cobertura de testes e facilidade de manutenção, usando métricas do site Code Climate⁴. O teste de usabilidade foi realizado por cinco 5 indivíduos especialistas na condução de RSL, sendo 3 na área de Psicologia e 2 em Economia, de forma individual, pelo tempo de uma hora, seguindo protocolo *Think-Aloud* (pensar em voz alta) [Fan et al. 2020]. Ao final, um questionário usando o *System Usability Scale* (SUS) [Brooke 1996] foi aplicado.

2.3. Métodos de Avaliações Encontrados

É comum que autores de novas tecnologias criem questionários de avaliação para serem preenchidos por usuários durante testes realizados com a tecnologia. Entretanto, quando não utilizam uma metodologia consolidada ou validada empiricamente, a avaliação conduzida pode não trazer uma interpretação sólida das respostas coletadas. Esta subseção destaca os métodos já consolidados e que foram usados pelos trabalhos citados anteriormente.

TAM, usada por StArt [Hernandes et al. 2012], é um dos modelos mais aplicados na área de Sistemas de Informação. Atua sob “a hipótese de que o uso de uma tecnologia da informação é determinado principalmente pela intenção de uso que o indivíduo apre-

³<https://www.cochrane.org/>

⁴<https://codeclimate.com/quality/>

sentada, e que essa intenção é definida pelas expectativas de melhoria esperada de desempenho decorrente do uso da tecnologia (utilidade percebida), e de facilidade de utilização da tecnologia (facilidade de uso percebida)” [de Brito and Ramos 2019].

Think-Aloud, usado em Buhos [Bustos Navarrete et al. 2018], é um protocolo onde o participante verbaliza seus pensamentos ao realizar tarefas [Fan et al. 2020]. É usado para obter percepções sobre o processo de pensamento do usuário que seriam difíceis de obter por mera observação [Fan et al. 2020].

SUS, também usado por Buhos, contém 10 perguntas pré-definidas com opções de respostas em escala Likert que ajudam a avaliar os componentes de efetividade (se os usuários conseguem completar seus objetivos), eficiência (quanto de esforço e recurso é necessário para realizar) e satisfação (em relação à experiência de uso). Ao final, uma pontuação (variando de 0 a 100) é calculada com base nas respostas fornecidas de acordo com cada pergunta, conforme orientação do método [Brooke 1996]. A interpretação da pontuação pode variar de acordo com cada autor. Bangor et al. [2009], por exemplo, afirmam que valores a partir de 70 indicam usabilidade aceitável.

Os métodos apresentados até aqui podem ser aplicados a diferentes tipos de solução. Thoth [Marchezan et al. 2019] e Buhos [Bustos Navarrete et al. 2018] usaram métodos específicos para avaliação de ferramentas de apoio para RSL, um proposto por Marshall et al. [2014] e outro por Al-Zubidy et al. [2019]. Estes métodos focam na verificação de cobertura de requisitos esperados em uma ferramenta de RSL. Os requisitos são agrupados e ao serem cobertos ou parcialmente cobertos recebem pontuação com associado peso de importância e, ao final, tem-se um resultado que permite comparar ferramentas. Estes métodos permitem ter uma visão geral de quão completa é a ferramenta e compará-la a outras que passarem pela mesma forma de avaliação.

O *framework* proposto por Marshall et al. [2014] aplica notas ponderadas⁵ para características cobertas, agrupadas aos conjuntos Economia, Facilidade de Introdução e Configuração, Apoio a Atividade de Revisão Sistemática e Gestão do Processo. O método de Al-Zubidy et al. [2019] agrupa os requisitos esperados em 11 grupos ordenados, os quais são: Apoio a Colaboração, Busca Integrada, Rastreabilidade, Apoio a *Text Mining*, Apoio a Inclusão e Exclusão, Apoio a Avaliação da Qualidade, Análise Automatizada, Visualização, Codificação de Métodos e Dados e Armazenamento de Estudos. Então, para cada requisito aplica uma nota ponderada⁶.

Na Tabela 1 é apresentado um resumo sobre os tipos e métodos de avaliação usadas nas ferramentas apresentadas. Percebe-se que a maioria dos trabalhos preocupou-se em combinar métodos quantitativos e qualitativos para realizar a avaliação da qualidade, o que entende-se também ser a melhor abordagem. Entretanto, nenhum trabalho seguiu um método de análise qualitativa específico sobre os dados coletados. Moller et al. [2015] apenas apresentou algumas frases para exemplificar alguns comentários recebidos de participantes. Thoth disponibilizou as questões do questionário para ser possível verificar

⁵Calculada quanto a importância: *Mandatory, High Desirable, Desirable, Nice to Have* e; grau de cobertura: completo, parcial ou nenhum

⁶Calculada quanto a ordem do grupo, quantidade de requisitos no grupo e grau de cobertura: total, parcial ou nenhum

Tabela 1. Métodos de Avaliação Utilizados na Ferramenta

Ferramenta	Quantitativa	Qualitativa
Buhos [Bustos Navarrete et al. 2018]	1. Score SUS; 2. Score AI-Zubidy; 3. Code-Climate	1. <i>Think-Aloud</i>
Rayyan [Ouzzani et al. 2016]	1. Precisão do algoritmo preditivo; 2. Questionário de avaliação não detalhado	1. Formulário de <i>feedback</i>
ReLiS [Bigendako and Syriani 2018]	Nenhuma	1. Avaliação qualitativa informal preliminar
SESRA [Molléri and Benitti 2015]	1. Questões em escala Likert; 2. Comparação de dados com amostras	1. <i>Feedback</i> de usuários
StArt [Hernandes et al. 2012]	1. TAM	1. Campos para comentários
Thoth [Marchezan et al. 2019]	1. <i>Score</i> adaptado de Marshall et al. [2014]; 2. Quasi-experimento com questões em escala Likert	1. Questões abertas, sem método definido

quais tipos de dados qualitativos foram coletados, mas nenhum resultado foi analisado.

3. Porifera

Porifera é uma aplicação *web* (porifera.app.br) voltada para pesquisadores que realizarão uma RSL. Foi desenvolvida sob acompanhamento de dois pesquisadores com experiência em diversas RSLs, os quais realizaram avaliação ad-hoc durante o processo. Algumas das sugestões propostas por eles foram: inclusão do cálculo do coeficiente Kappa, evitar a visualização das revisões de outros pesquisadores antes de fazer sua própria revisão, uso de estratégia para composição da *string* de busca, entre outras.

Na Porifera, após se cadastrar, cada pesquisador pode criar mais de um projeto de RSL. Cada projeto deve possuir um nome. Ao criar um projeto o pesquisador torna-se proprietário do projeto e sob esta condição pode realizar algumas atividades, como definir o planejamento, configurar aspectos da condução, importar publicações e exportar dados do projeto. O proprietário também pode convidar outros pesquisadores para serem co-autores. Estes novos pesquisadores poderão assumir papéis como colaborador, co-orientador e orientador.

3.1. Planejamento da RSL

Na definição do planejamento, o proprietário pode indicar: objetivos, estratégia de pesquisa, fontes de pesquisa e critérios para seleção dos estudos primários. O objetivo pode ser definido com ajuda de uma frase modelo baseado no método *Goal-Question-Metric (GQM)* [Basili and Caldiera 2000]. Além do objetivo, é possível também registrar a questão de pesquisa principal. Com relação à estratégia de pesquisa, Porifera permite registrar a *string* de busca base que será usada para pesquisa. Opcionalmente, pode-se usar a estrutura *Population, Intervention, Comparison, Outcome, Context (PICOC)* [Petticrew and Roberts 2005] para auxiliar na definição da *string* ou ainda recorrer a uma busca por sinônimos e palavras relacionadas. Quanto às fontes da RSL, deve-se adicionar as bibliotecas digitais que serão consultadas. Para cada biblioteca é possível especificar a *string* de busca para atender a peculiaridades do buscador da biblioteca digital. Para algumas delas como IEEExplore⁷ e Scopus⁸. Para as atuais versões destas bibliotecas digitais, até a data de escrita deste artigo, Porifera faz a adaptação da *string* automaticamente. Por meio de um botão na interface é possível também abrir uma janela da biblioteca e testar a aplicação da *string* definida. Por fim, os critérios de inclusão e exclusão que serão usados no processo de seleção de estudos devem ser também definidos. Para cada critério define-se um código, uma descrição e um mnemônico, que é uma palavra que facilitará lembrar rapidamente o objetivo do critério quando apresentada em alguns controles de interface.

As publicações são carregadas por importação de arquivos no formato Bibtex⁹. Estes arquivos são obtidos pela exportação nas bibliotecas digitais dos resultados da pesquisa obtida pela *string* de busca definida. Porifera registra a data que cada publicação foi importada e associa a publicação à respectiva biblioteca digital. Porifera também permite realizar importações futuras da mesma biblioteca, possibilitando carregamentos parciais ou mesmo atualizar a RSL a qualquer tempo, com novas publicações sem provocar duplicadas da mesma biblioteca e sem interferir nas avaliações já realizadas em publicações já existentes. O processo de importação também exibe registros com dados incompletos que merecem atenção do pesquisador, permitindo analisar, corrigir e tomar a melhor decisão durante a importação.

A lista com as informações das publicações importadas em Porifera (Figura 1) possui um painel de controle de listagem, por meio do qual pode-se customizar as publicações exibidas, permitindo observar diferentes facetas do conjunto de publicações e navegar rapidamente entre elas. O controle permite aplicar filtros (Figura 1, item 1) de forma individual ou combinada, para, por exemplo: listar somente as publicações que ainda não foram revisadas pelo pesquisador logado; listar somente as publicações com acesso aberto; listar somente as publicações de uma biblioteca digital específica e; listar as publicações obtidas de uma biblioteca específica, aceitas pelo grupo e cujo *link* para arquivo ainda não foi associado. Além dos filtros, é possível pesquisar por publicações no campo autor, título, resumo e *Digital Object Identifier (DOI)* (Figura 1, item 2). A listagem pode ser apresentada em diferentes tamanhos de paginação e ordenadas por número, título, autores e ano (Figura 1, item 3).

⁷<https://ieeexplore.ieee.org/Xplore/>

⁸<https://scopus.com/>

⁹<http://www.bibtex.org/Format/>

Review Publications

for protocol criteria activities

Phase Progress 10

First Review 14%
3006 publications still awaiting decision

Second Review 10%
78 publications still awaiting decision

Your Progress 11

First Review 10%
2931 publications still awaiting your review

Second Review 17%
72 publications still awaiting your review

Final Consensus 12

First Review 57% accepted
438 rejected

Second Review 2 accepted
7 rejected

Agreement and reliability more... 13

First Review
Fleiss Kappa: 0.7991 / Agreement: 92.95%
About 525 publications.

Second Review
Fleiss Kappa: 0.2286 / Agreement: 55.56%
About 9 publications.

Filter By 1

Source: all

Phase: any

Review: all publications

Flags: indifferent

Duplicate: all publications

Criteria: all

Decision: indifferent

File Link: all publications

Search 2

By: put a word or #number here

in authors title abstract

number (if numeric)

Go! clear search

Show

Other's review: Showing

Publications/page: 10

Order by: number

List Total of 3531 publications in this project
Displaying page 1 of 354.

Publication Info 4

#1, Collected at 23/02/2021 from Scopus

Classified by source as: **Conference Paper**

Xeferis S., Palaigeorgiou G., Zoumpourtikoudi H., 2021

Educational Robotics for Creating "Tangible Simulations": A Mixed Reality Space for Learning the Day/Night Cycle

▶ abstract

6 this seems be unique on project!

5 open cite this google it! translate open file 8

Reviews 7

First Second

Yours

11 Parece conter uma tecnologia tocável. Precisa verificar detalhes

By Others

11 :

16 : Não aborda a avaliação de uma solução holográfica

Is there a consensus in the reviews? has no consensus

Final Consensus Decision

Publication accepted by criteria 11: TOMCH HOLO/RA! Undo

Consensus registered at 2021-02-16 15:09:32 by

Publication passed to next phase!

Figura 1. Parte da tela de Porifera contendo a lista de publicações de uma RSL.

3.2. Processo de Seleção

A seleção dos estudos primários ocorre sobre a lista de publicações. Ao lado (Figura 1, item 7) da informação de cada publicação encontra-se o espaço destinado à aplicação dos critérios de inclusão/exclusão. O processo de seleção tem um fluxo bem definido e acontece em duas fases (filtros), identificadas pelas abas “First” e “Second”. De forma prática, o pesquisador seleciona um critério, listado por meio de código e mnemônico e, opcionalmente, acrescenta um comentário sobre sua avaliação. Intencionalmente, o pesquisador não enxerga as revisões de seus colegas enquanto não realizar sua própria avaliação sobre o estudo. Quando há concordância dos critérios nas avaliações, o sistema avisará o proprietário do projeto e este poderá confirmar o fato. Se houver um desacordo entre a opinião dos pesquisadores, o sistema avisa haver discordância entre as avaliações. Os pesquisadores podem discutir sobre a publicação e ao final, o proprietário do projeto poderá sinalizar um consenso decidido pelos pesquisadores envolvidos. A atividade de atribuir critérios às publicações pode ser realizada de forma assíncrona, cada pesquisador ao seu tempo. A reunião de discussão pode ocorrer de forma presencial ou por meio de tecnologia de videoconferência, ou *chat*. As decisões já tomadas para cada publicação são apresentadas indicando se foi por concordância ou consenso, para fins de rastreabilidade e as avaliações de cada pesquisador mantidas para fins de auditoria. Estas funcionalidades, com este fluxo de interação, não foram encontradas nas ferramentas relacionadas.

Além da aplicação dos critérios, o sistema traz outras facilidades úteis ao processo de seleção de estudos. Permite sinalizar (Figura 1, item 8) cada publicação com um símbolo de estrela, indicando uma certa relevância/importância daquela publicação no contexto da pesquisa ou com o símbolo de um par de óculos, indicando que se deve

olhar mais profundamente sobre aquela publicação na próxima fase da seleção a fim de confirmar as suspeitas de utilidade dela para a pesquisa. Para as publicações aceitas para a fase seguinte, também é possível adicionar um *link* para o documento da publicação (Figura 1, item 9), que pode estar armazenado em qualquer serviço de nuvem de arquivos que gere URL de acesso. Deste modo, as publicações podem ser facilmente acessadas pelo grupo de pesquisadores e, inclusive, manter centralizado a adição de comentários e anotações sobre estes documentos.

Porifera apresenta durante o processo de seleção de estudos um painel para acompanhamento de algumas informações sobre cada fase da seleção. Compõem o painel: o índice de publicações já avaliadas pelo grupo (Figura 1, item 10), e individualmente pelo pesquisador autenticado (Figura 1, item 11); a quantidade de estudos aceitos e rejeitados (Figura 1, item 12); o índice de concordância e um índice de confiabilidade em relação às avaliações (Figura 1, item 13) - o Kappa Cohen [Silva and Paes 2012] e o Kappa Fleiss [Fleiss 1971]. Porifera calcula automaticamente e apresenta o índice Kappa Cohen para projetos com até dois pesquisadores e o índice Kappa Fleiss para projetos com 3 ou mais pesquisadores. Para projetos de 3 pesquisadores o sistema também apresenta o índice Kappa Cohen para cada dupla de pesquisadores, o que facilita identificar maiores desníveis de critérios entre o grupo, permitindo uma ação para nivelamento. A ferramenta apresenta a interpretação dos valores de Kappa entre 0 e 1, conforme proposto por diferentes autores, permitindo que o grupo escolha qual interpretação irá utilizar.

Para atender a necessidade de ajuda ao usuário, Porifera oferece alguns recursos como: ajuda contextual (*inline*) e uma página de perguntas frequentes (*Frequently Asked Questions, FAQ*) e um tutorial contendo texto e imagens demonstrando como realizar passo a passo as principais operações na ferramenta. Algumas tarefas, como o processo de importação, são realizadas por meio de Assistente de Execução (*wizard*). Porifera também possui uma política de uso e privacidade para informar diferentes aspectos do seu funcionamento ao usuário, bem como se adequar à legislação.

4. Avaliação de Porifera

Porifera foi disponibilizada para uso e avaliação da sua qualidade durante a realização de uma atividade proposta em disciplina com foco em pesquisa experimental nas áreas de ES e IHC, do Programa de Pós-Graduação em Informática da UFPR. O objetivo foi obter a percepção dos estudantes em relação à aceitação de Porifera considerando os aspectos: facilidade de uso, utilidade, intenção de uso futuro e eficácia. A avaliação também objetivou identificar as emoções dos usuários no decorrer do uso de Porifera, pontos de melhoria e novos recursos que pudessem apoiar a execução de uma RSL colaborativa(o).

4.1. Participantes

Participaram do estudo 11 alunos (identificados de P01 a P11). Todos autorizaram o uso de suas respostas de forma anonimizada nesta avaliação para fins de pesquisa e assinaram um Termo de Consentimento Livre e Esclarecido (TCLE). O perfil de cada participante está apresentado na Tabela 2. Dos 11 alunos, 9 são de pós-graduação em Informática (5 de mestrado e 4 de doutorado), 1 de graduação em Ciência da Computação e 1 de Informática Biomédica. Destes, seis são mulheres e cinco são homens. Cinco pessoas

Tabela 2. Perfil dos Participantes

Id	S	Faixa Etária	Curso	Experiência em RSL	Ferramentas	Colab
P01	F	36-40	Doutorado	Intermediário	Parsifal, StArt	Sim
P02	F	36-40	Mestrado	Iniciante	ReLiS, StArt	Não
P03	M	26-30	Doutorado	Intermediário	Parsifal	Sim
P04	F	20-25	Mestrado	Iniciante	—	Sim
P05	F	20-25	Mestrado	Iniciante	—	Não
P06	F	46-50	Doutorado	Iniciante	—	Sim
P07	M	20-25	Graduação	Iniciante	—	Não
P08	M	41-45	Mestrado	Iniciante	—	Não
P09	M	41-45	Doutorado	Iniciante	—	Não
P10	M	20-25	Mestrado	Iniciante	—	Não
P11	F	20-25	Graduação	Iniciante	—	Não

Legenda:

Id - Participante; **S** - Sexo (F, Feminino; M, Masculino); **FE** - Faixa Etária; **Ferramentas** - Experiência com Ferramentas de apoio para RSL; **Colab** - Se já realizou uma RSL com outros pesquisadores.

com idade entre 20 e 25 anos; uma entre 26 e 30 anos; duas entre 36 e 40 anos; duas entre 41 e 45 anos e; uma entre 46 e 50 anos.

Nove participantes disseram terem conhecimento inicial sobre RSL, começando a aprender sobre o tema ou já aprenderam sobre o assunto, mas realizaram o processo apenas uma vez e; dois afirmaram ter conhecimento intermediário, tendo realizado uma RSL mais de uma vez. Ninguém indicou ter conhecimento avançado ou que realiza o processo de RSL com frequência. Poucos participantes relataram conhecer uma ferramenta de apoio específico para RSL, citadas Parsifal e StArt [Hernandes et al. 2012] (2 vezes cada) e ReLiS (uma vez). A maioria (7) nunca realizou uma RSL com outros pesquisadores, enquanto 4 já haviam realizado.

4.2. Metodologia

Para avaliar Porifera foram misturados métodos de análise quantitativa e de análise qualitativa. Foi escolhido o método TAM, que teve seu formulário original adaptado, acrescido sentenças relacionadas à eficácia e algumas questões abertas. A partir deste formulário foram realizadas análise quantitativa e qualitativa. O mesmo tipo de análise ocorreu sobre um formulário de *feedback* com campos de escopo determinado e questão aberta. O TAM foi escolhido para obter por meio de um modelo consolidado diferentes percepções em relação à ferramenta. Primeiro, sobre o aspecto da facilidade de uso da ferramenta como interface e acesso aos recursos necessários para a condução de uma RSL colaborativa. Em segundo sobre o aspecto da utilidade, para garantir que Porifera traga uma mínima contribuição ao processo de RSL. E, em terceiro, sobre o aspecto da intenção de uso, para avaliar se seria interessante continuar investindo esforços de projeto e implementação em uma nova ferramenta para RSL.

Por entender que Porifera é uma ferramenta ainda em estágio inicial e que algumas

funcionalidades desejáveis ainda não estão implementadas (há funcionalidades em *backlog* para futuras versões), estes dois métodos foram escolhidos para avaliar a intenção de uso, usabilidade e consequente viabilidade do prosseguimento do projeto. Os métodos com foco na cobertura de requisitos, propostos por Marshall et al. [2014] e Al-Zubidy et al. [2019], não seriam, agora, adequados para avaliar Porifera, visto que muitas funcionalidades não estão implementadas.

Esta avaliação pode ser classificada como exploratória e descritiva. O uso de Porifera foi dividido em duas etapas. Na etapa 1, Porifera foi apresentada aos estudantes durante 30 minutos em uma aula remota síncrona por meio de videoconferência. Nesta apresentação, foi demonstrado como realizar o cadastro; como criar um projeto e definir aspectos do planejamento da RSL, tais como objetivos, questões de pesquisa, *string* de busca, bibliotecas digitais a serem usadas e critérios de inclusão e exclusão e; como definir papéis e convidar pesquisadores para colaborar com a RSL. Aos alunos foi dada a atividade de se cadastrar na ferramenta e criar um projeto de RSL, conforme seu interesse de pesquisa atual. Eles também deveriam convidar duas pesquisadoras pré-definidas (monitoras da disciplina) para serem colaboradoras da RSL. Os alunos tiveram uma semana para realizar esta atividade.

Na etapa 2, em uma segunda aula, durante uma apresentação de 30 minutos, foi demonstrado como realizar a importação de publicações de uma biblioteca digital, como navegar pela lista de publicações e como realizar a seleção dos estudos aplicando os critérios em duas fases (filtros) - a primeira baseando-se apenas em metadados da publicação e a segunda baseando-se na leitura completa da publicação. Também foi apresentado como acompanhar o progresso da revisão e verificar o nível de concordância e confiabilidade. Os alunos tiveram também uma semana para importar pelo menos 30 publicações resultantes de sua pesquisa em uma biblioteca digital em Porifera. Além disso, foi solicitado que eles aplicassem seus critérios nas 30 publicações importadas no 1º filtro. Em seguida, as colaboradoras também realizaram a avaliação destas 30 publicações e indicaram seus critérios, conforme sua percepção. Depois disso, os participantes da pesquisa, como proprietários do projeto, deveriam verificar se houve concordância ou desacordo na definição dos critérios em cada filtro, confirmando as concordâncias ou moderando a decisão consensual sobre um critério de inclusão/exclusão. O processo se repetiu da mesma forma durante o segundo filtro, desta vez apenas sobre as publicações aceitas no primeiro filtro.

Nas duas etapas, caso houvessem dificuldades, os alunos poderiam recorrer a um tutorial *online* de Porifera e a um e-mail de suporte. Destaca-se que o tempo de duas semanas para executar estas atividades não são suficientes para realização de uma RSL. Porém, serviu para introduzir ao aluno o contexto de planejamento e condução de uma RSL, bem como ao uso de uma ferramenta de apoio para RSL. A professora sugeriu que os alunos deveriam revisar o planejamento e continuar o processo de RSL referente à sua pesquisa em conjunto com seu orientador.

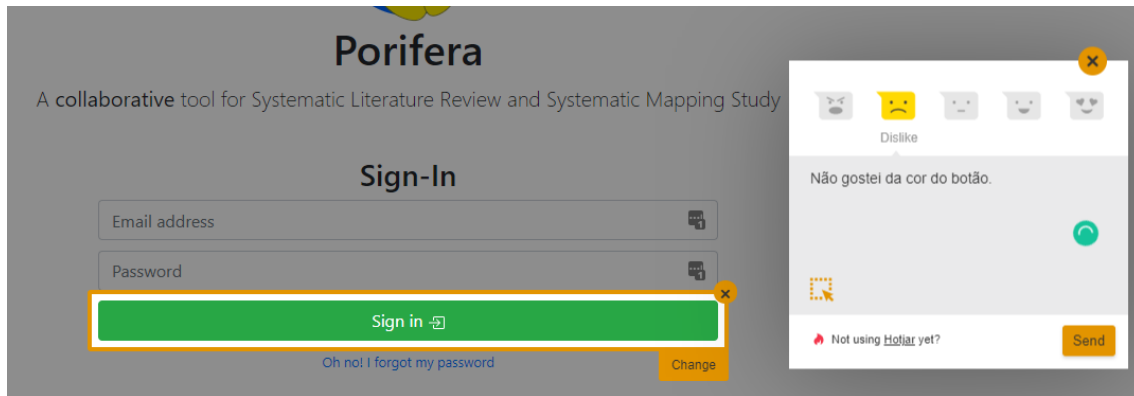


Figura 2. Parte da tela de Porifera quando recurso para dar *feedback* está ativo

4.3. Procedimento de Coleta de Dados

Porifera possui um recurso integrado de *feedback*, proporcionado por uma solução externa chamada Hotjar¹⁰. Por ele, em qualquer tela da aplicação o usuário pode enviar um *feedback*, escolhendo um *emoji* que melhor representa sua emoção no momento (Figura 2). Hotjar registra de forma anonimizada a URL onde ocorreu o *feedback* e guarda um *screenshot* da tela (Figura 3). Opcionalmente, o usuário pode adicionar um comentário e sinalizar uma área da tela a qual está se referindo. Os *feedbacks* são registrados de forma anônima e ficam disponíveis para consulta dos responsáveis por Porifera. O recurso permite envio de ilimitados *feedbacks* por usuário. O agrupamento dos *feedbacks* por usuário considera *cookies* de navegador. Durante o período de avaliação de Porifera, 24 *feedbacks* foram recebidos¹¹, sendo 20 com comentários e agrupados em 9 diferentes usuários anonimizados.

Além da coleta de *feedback*, ao final da experimentação os participantes responderam um questionário de avaliação onde puderam relatar os principais problemas e dificuldades encontrados ao realizar as atividades propostas, bem como indicar suas percepções quanto a facilidade, utilidade, eficácia e intenção de uso futuro de Porifera.

O questionário continha sentenças para serem respondidas em escala Likert¹² para cada um dos componentes do TAM (Tabela 3, Tabela 4 e Tabela 5). Também trazia questões objetivas para verificar a eficácia de alguns dos aspectos de Porifera (Tabela 6), seguindo a mesma escala Likert. Para coletar sugestões e percepções, quatro questões abertas foram feitas (Tabela 7).

5. RESULTADOS

As emoções indicadas pelo formulário de *feedback* e as questões em escala Likert do questionário final foram analisadas em relação à distribuição da resposta e a frequência (moda e mediana) das opções em cada uma das sentenças. Deste modo, foi possível ter uma

¹⁰<https://www.hotjar.com>

¹¹Planilha com os dados estará disponível no Figshare.

¹²Concordo Totalmente (CT); Concordo Parcialmente (CP); Não concordo; Nem discordo (NN); Discordo Parcialmente (DP) e; Discordo Totalmente (DT)

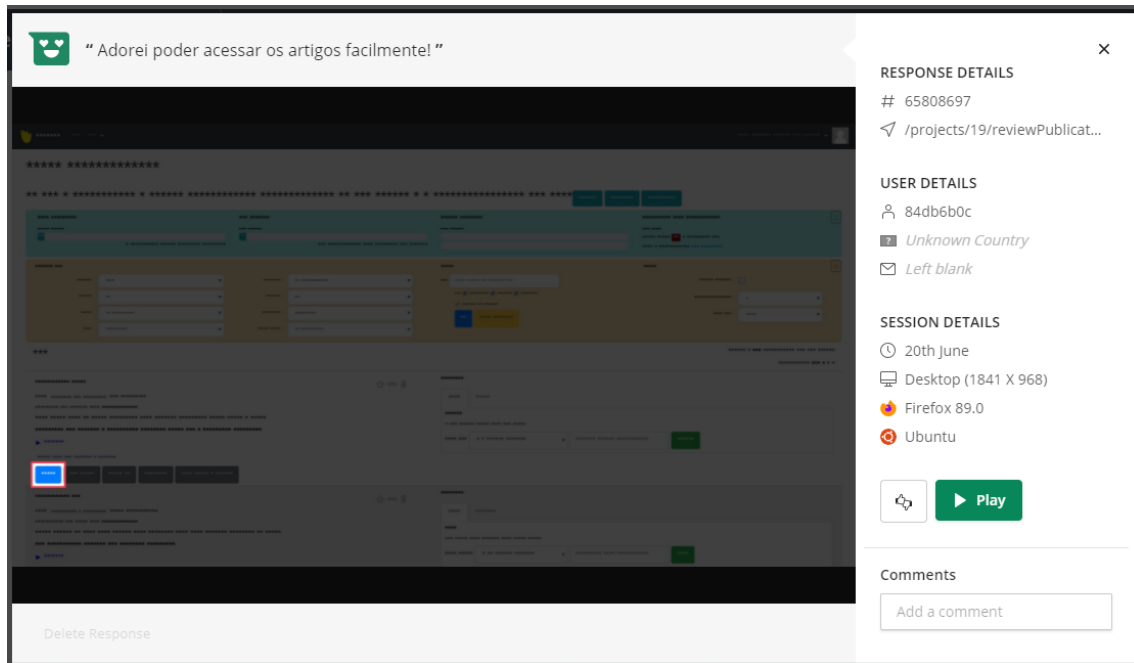


Figura 3. Exemplo de informações registradas por meio de Hotjar.

Tabela 3. Sentenças relacionadas a Facilidade de Uso

Id	Sentença
FU1	Minha interação com Porifera foi clara e compreensível
FU2	Interagir com Porifera exige pouco esforço mental
FU3	Considero Porifera fácil de usar
FU4	Considero fácil utilizar Porifera para fazer o que quero que ela faça, isto é, apoiar atividades de uma RSL.

Tabela 4. Sentenças relacionadas a Utilidade Percebida

Id	Sentença
UP1	Utilizar Porifera melhorou o meu desempenho na realização de uma RSL em grupo
UP2	Utilizar Porifera permitiu aumentar minha produtividade na realização de uma RSL em grupo
UP3	Usar Porifera aumentou minha eficácia na realização de uma RSL em grupo
UP4	Considero Porifera útil para realização de uma RSL em grupo.

Tabela 5. Sentenças relacionadas a Intenção de Uso Futuro

Id	Sentença
IUF1	Supondo que eu tenho acesso à Porifera, eu pretendo usá-la.
IUF2	Levando em conta que eu tenha acesso a Porifera, eu prevejo que iria usá-la.

Tabela 6. Sentenças relacionadas a Eficácia

Id	Sentença
E1	A importação de dados e o processo de revisão dos estudos realizado por meio de Porifera previne a ocorrência de erros que poderiam acontecer em atividades manuais.
E2	O processo de resolução de desacordos ou confirmação na revisão funcionam como se espera em uma RSL realizada(o) em grupo.
E3	Os recursos oferecidos por Porifera atenderam minhas expectativas em relação a recursos que eu esperava encontrar em uma ferramenta de apoio para RSL.
E4	A forma de solução proposta (aplicação <i>web</i> para uso em <i>desktop/notebook</i>) é adequada para a realização de uma RSL em grupo.

Tabela 7. Questões Abertas

Id	Sentença
QA1	Comente sobre as dificuldades de uso que você identificou ao experimentar Porifera.
QA2	Comente sobre a utilidade ou não compreensão de utilidade da Porifera e alguns de seus recursos.
QA3	Comente sobre sua perspectiva de uso futuro de Porifera.
QA4	Indique algumas sugestões para melhoria de Porifera. Pontos dos já existentes que podem ser melhorados ou recursos que desejaria encontrar ou ter a disposição na ferramenta.

percepção do sentimento geral de uso de Porifera e uma compreensão da concordância dos participantes sobre cada sentença. Os resultados estão descritos na Subseção 5.1.

Sobre as respostas das questões abertas e sobre os comentários enviados junto ao *feedback* foi realizada uma análise qualitativa usando um subconjunto de procedimentos do método *Grounded Theory* (GT) [Corbin and Strauss 2014]. Os comentários extraídos dos questionários e os recebidos por meio da ferramenta de *feedback* foram analisados separadamente, em razão da natureza da identificação da coleta e também para que se pudesse realizar a triangulação de dados.

Primeiro os comentários foram extraídos dos questionários e colocados em documentos de texto, um para cada participante. Depois os documentos foram analisados por codificação aberta (1a etapa do GT) e agrupado de acordo com às suas semelhanças e características que finalmente poderiam representar conceitos de categorias e subcategorias (codificação axial - 2a fase do GT). A codificação seletiva (3a fase do GT) não foi realizada, pois não se pretendia criar uma teoria. As etapas de codificação aberta e axial foram suficientes para compreender as dificuldades enfrentadas pelos participantes ao utilizar Porifera. Os resultados qualitativos são apresentados nas subseções 5.2 e 5.3.

5.1. Análise Quantitativa

Com relação à emoção indicada no *feedback*, utilizando a classificação do Hotjar tivemos: 10 *Love* (41,7%), 3 *Like* (12,5%), 2 *Neutral* (8,3%), 9 *Dislike* (37,5%) e nenhum *Hate*. A emoção com maior frequência foi *Love* e, se converter as opções em valores numéricos

de 1 a 5, a mediana seria 4, correspondendo a *Like*. Fazendo um recorte para os *feedbacks* onde foi sinalizada a emoção *Dislike* verificou-se que os comentários reportam erros da ferramenta. Alguns exemplos de avaliações com *Love* ocorreram na página de listagem de publicações, com comentários como “Adorei poder acessar os artigos facilmente!” e “Fantástico poder visualizar os dados tão facilmente!”. Como exemplos de avaliações com *Dislike* podemos citar interações na página de edição do projeto, com comentários que indicavam falhas no funcionamento das telas: “Estava dando alguns erros ao tentar adicionar uma regra de Excluído/Reject...” e “Eu selecionei o GQM em português (...) aí voltei para mudar o GQM para o inglês (...) e perdi o que já havia salvo. O dislike é só pela situação relatada, porque achei muito bom o sistema que estão desenvolvendo”.

Com relação ao local que motivou os *feedbacks*, 1 foi dado na página inicial, antes de o usuário fazer *logon*, 4 na página inicial após o *logon* do usuário, 1 na tela de cadastro, 4 na tela de definição do planejamento da RSL, 1 na tela de listagem e convite de pesquisadores ao projeto, 1 na tela de importação de publicações e a maioria, 12, na tela de listagem, usada também para seleção de estudos.

Para todas as sentenças do TAM, a moda e a mediana foram a opção “concordo totalmente”. Em outras palavras, “concordo totalmente” foi a resposta mais frequente em todas as sentenças. Não existiu nenhuma resposta com a opção “discordo totalmente”.

Em relação à Facilidade de Uso (Figura 4), um dos participantes discordou parcialmente e outro não concordou nem discordou na sentença FU2. Isto pode indicar que a ferramenta ainda precisa de ajustes a fim de exigir menos esforço mental.

A soma das porcentagens de respostas de “concordo totalmente” (CT) com “concordo parcialmente” (CP) variaram de 81,82% na menor frequência (para FU2) a 100% (para FU1, FU3, FU4, UP1, UP2, UP3, UP4, IUF1 e IUF2). Ou seja, a maioria dos participantes concorda (total ou parcialmente) que Porifera é fácil de usar, útil e têm intenção de usá-la.

Com relação às sentenças sobre a eficácia de Porifera, observa-se (Figura 5) total concordância quanto ao fato que é adequado usar uma aplicação *web* para apoiar a realização de uma RSL em grupo (E4). Além disso, há elevado nível de concordância de que os recursos oferecidos por Porifera atenderam as expectativas em relação a recursos esperados em ferramentas de apoio para RSL (E3). Neste mesmo sentido, os participantes concordaram que os recursos de importação de dados e o processo de revisão realizado por meio de Porifera previne a ocorrência de erros (E1) e; que o processo de resolução de desacordos ou confirmação da concordância nas respostas funcionam em Porifera como se esperava encontrar em uma ferramenta colaborativa de apoio a RSL (E2). No entanto, o participante P9 não concordou nem discordou das sentenças E1, E2 e E3, indicando que os recursos de Porifera podem ter sua eficácia melhorada.

5.2. Análise Qualitativa das Questões Abertas no Questionário

O agrupamento dos códigos obtidos fizeram emergir seis categorias: Percepções de Facilidade, Percepções de Utilidade, Percepções de Intenção de Uso, Dificuldades no Uso, Sugestões de Melhoria e Emoções Expressadas Durante o Uso.

Facilidade de Uso											CT+CP	
FU1	P3	P7	P10	P1	P2	P4	P5	P6	P9	P9	P11	100,00%
FU2	P3	P7	P4	P6	P11	P1	P2	P5	P9	P9	P10	81,82%
FU3	P3	P6	P7	P9	P10	P1	P2	P4	P5	P9	P11	100,00%
FU4	P9	P10	P1	P2	P3	P4	P5	P6	P7	P9	P11	100,00%

Utilidade Percebida												
UP1	P9	P1	P2	P3	P4	P5	P6	P7	P9	P10	P11	100,00%
UP2	P1	P2	P3	P4	P5	P6	P7	P9	P9	P10	P11	100,00%
UP3	P6	P9	P1	P2	P3	P4	P5	P7	P9	P10	P11	100,00%
UP4	P1	P2	P3	P4	P5	P6	P7	P9	P9	P10	P11	100,00%

Intenção de Uso Futuro												
IUF1	P1	P2	P3	P4	P5	P6	P7	P9	P9	P10	P11	100,00%
IUF2	P10	P1	P2	P3	P4	P5	P6	P7	P9	P9	P11	100,00%

DT	DP	NN	CP	CT
----	----	----	----	----

Figura 4. Distribuição de respostas por sentenças do TAM

Eficácia												
E1	P9	P1	P2	P3	P4	P5	P6	P7	P9	P10	P11	90,91%
E2	P9	P3	P6	P7	P1	P2	P4	P5	P9	P10	P11	90,91%
E3	P9	P1	P2	P3	P4	P5	P6	P7	P9	P10	P11	90,91%
E4	P1	P2	P3	P4	P5	P6	P7	P9	P9	P10	P11	100,00%

DT	DP	NN	CP	CT
----	----	----	----	----

Figura 5. Distribuição de respostas relacionadas a Eficácia

5.2.1. Percepções de Facilidade

Porifera foi considerada intuitiva (ver citação de P1 abaixo), prática e fácil de usar (ver citação de P8). Um dos participantes indicou que os recursos estão claros para uso (ver citação de P9) e outro, que Porifera possui uma interface visual de fácil entendimento (ver citação de P10).

P1 *“Não houve dificuldades, a ferramenta é bastante intuitiva”*

P8 *“Pretendo utilizar a ferramenta na minha pesquisa, pois é prática e fácil de usar”*

P9 *“A princípio os recursos estão bem claros para utilização.”*

P10 *“(...) principalmente a parte visual de fácil entendimento sobre os filtros, índices de confiança e etc”*

5.2.2. Percepções de Utilidade

Porifera é útil de modo geral (ver citação de P06 abaixo) e possui recursos úteis para gerenciar aplicações (ver citação de P8). Ao usar Porifera, a realização de uma RSL colaborativa(o) torna-se mais ágil (ver citação de P01), destacando-se os filtros que realizam a seleção de estudos em duas fases (ver citação de P4). Também foi considerado um facilitador a interface que ajuda a construir a *string* de busca, a avaliação de publicações por aplicação de critérios pré-definidos e os filtros que controlam a lista de publicações exibidas (ver citação de P5). Em geral, percebe-se que Porifera facilita a execução de uma RSL (ver citação de P2).

P6 *“Sobre a utilidade da Porifera, é fantástica”*

P8 *“(...) os recursos são úteis para o gerenciamento das publicações.”*

P1 *“A possibilidade de utilizar Porifera de forma colaborativa agiliza a execução das atividades da RSL”*

P4 *“Com a Porifera ocorreu a agilidade no processo de filtros para a RSL que foi de grande importância.”*

P5 *“(...) facilitou muito o processo principalmente na parte de criação da string de busca que já formatava com base na biblioteca e também na avaliação dos artigos com base nos critérios e todos os filtros que podem ser aplicados para facilitar a busca nesta etapa.”*

P2 *“Porifera é uma ótima ferramenta que facilita muito a RSL em todas as fases.”*

5.2.3. Percepções de Intenção de Uso

Muitos participantes revelaram que pretendem continuar usando Porifera em suas pesquisas (ver citação de P5), ou considerá-la com seus pares (veja citação de P3). Um dos participantes até recomendou Porifera à outra pesquisadora (ver citação de P7).

P5 “(...) *pretendo utilizar a Porifera para fazer uma RSL melhor para a minha pesquisa (...)*”

P3 “*Pretendo discutir o uso da ferramenta com meus professores para dar continuidade à pesquisa.*”

P7 “*Gostei a ponto de recomendar para a minha namorada que cursa Direito e a mesma ficou interessada.*”

5.2.4. Dificuldades no Uso

Como também se esperava, a avaliação identificou algumas dificuldades presentes em Porifera. Os participantes relataram principalmente duas dificuldades. Primeiro, a de lembrar como visualizar os critérios aplicados pelos outros pesquisadores do projeto (ver citação de P8 abaixo). Isto porque, o elemento de interface que ativa esta visualização vem desativado por padrão (ver citação de P2) sendo desenhado de forma muito discreta, dificultando lembrar ou fazer-se entender ser necessário ativá-lo para que a classificação dada pelos colegas fosse exibida. Motivo pelo qual alguns participantes precisaram de ajuda, revendo a apresentação da ferramenta (ver citação de P3 abaixo). Este *feedback* motivou um redesenho do componente (Figura 6).

P8 “*Apenas uma dificuldade pontual em relação à visualização das revisões efetuadas pelas colaboradoras. Não havia marcado a caixa de seleção da opção de visualização.*”

P2 “*Na parte de decisões por padrão vem desabilitado e não fica claro que precisa clicar no item para aparecer as outras decisões.*”

P3 “*Tive que rever a parte da demonstração que mostrava a parte de consenso. Pois antes de sinalizar consenso, era necessário marcar o checkbox do filtro de “Other’s review”. (...) não tinha lembrado.*”

A segunda dificuldade foi identificada na tela de definição de critérios de inclusão e exclusão. Houve participante que confundiu as colunas dos critérios de exclusão com os de inclusão (ver citação de P5a abaixo); houve quem não conseguisse compreender os campos obrigatórios para preencher durante a criação do critério (ver citação de P2). Teve ainda, participantes que tentaram alterar um critério já existente, mas tal funcionalidade ainda não estava implementada e o botão estava desativado (ver citação de P9). Dado estes *feedbacks* esta tela deverá ser revisada para assegurar que os usuários consigam compreender e realizar todas as operações relacionadas a critérios de inclusão e exclusão.



Figura 6. Elemento de interface modificado após avaliação

P5a *“Ao adicionar os critérios de inclusão e exclusão, eu pensei que do lado esquerdo seriam os critérios de inclusão e do direito os critérios de exclusão, porque geralmente ações “negativas” ficam do lado direito. Mas depois de adicionar todos os critérios de inclusão e iniciar os de exclusão que percebi que havia feito errado”*

P2 *“Na parte de Critério em alguns monitores esse “cinza”, não fica cinza, fica parecendo que já está digitado a informação, o que confunde o usuário se precisa incluir ou não o dado.”*

P9 *“Acho que ainda falta poder editar os critérios de uso, que apesar de não ser uma dificuldade propriamente dita, pode tornar cansativo excluir e depois adicionar novamente.”*

5.2.5. Sugestões para Melhoria

Aqui foram reunidos os comentários que demonstram alguma necessidade não coberta por Porifera, como novas funcionalidades ou explicitam sugestão de melhoria na interface dada pelos participantes (ver citação de P3, P8 e P11 abaixo). Dentre as funcionalidades sugeridas, pode-se citar como exemplo: exportar dados em formato de planilha eletrônica, editar dados das publicações importadas, incluir *chat* para comunicação entre pesquisadores durante discussões de consenso, permitir avaliar as publicações produzindo ranking de qualidade, indicar importação do tipo *snowballing* e automatização da aplicação de critério de exclusão para trabalhos duplicados.

P3 *“Uma sugestão talvez seja a criação automática do critério de exclusão para trabalhos duplicados. Quando identificados que sejam rotulados ou de fato removidos. ”*

P8 *“algumas melhorias na usabilidade, botão claro de visualização de revisões”*

P11 *“(…) creio que aprimorar o design da plataforma pode fazer com que o ambiente pareça mais amigável”*

5.2.6. Emoções Expressadas Durante o Uso

Por fim, a análise qualitativa das questões abertas permitiu também identificar que Porifera despertou emoções positivas nos participantes. De uma forma geral, os participantes gostaram de usar a ferramenta (ver citação de P5) e ficaram surpresos pelas funcionalidades e utilidades apresentadas (ver comentário de P11 abaixo).

P5 *“Gostei muito de usar o Porifera”*

P11 *“Me surpreendi com as funcionalidades e utilidades apresentadas na Porifera”*

5.3. Análise Qualitativa dos *Feedbacks*

O agrupamento dos códigos obtidos a partir dos *feedbacks* anônimos fizeram surgir quatro categorias: “Relatos de Erro”, “Oportunidades de Melhorar Usabilidade” “Sugestões” e “Utilidade e Facilidade de Uso”. Nesta seção, os participantes e seus comentários serão identificados pela sigla “PA” (Participante Anônimo), uma vez que não há forma de relacionar os *feedbacks* obtidos com o formulário TAM preenchido pelos participantes. Ou seja, PA1 não é a mesma pessoa que P1, embora o conjunto de PA’s e P’s correspondam aos mesmos participantes do estudo.

5.3.1. Relatos de Erro

Os participantes relataram erros durante o uso de algumas funções da ferramenta. Um mesmo participante (veja as citações PA04a, PA04b e PA04c abaixo) relatou três erros: durante o cadastro, ao tentar realizar ordenação de listagem e o não tratamento adequado de alguns caracteres especiais durante a importação. Outros erros reportados por outros usuários ocorreram ao tentar criar critério de exclusão durante o planejamento (ver citação do participante PA03 abaixo) e ao tentar aplicar um filtro na listagem escolhendo um critério (ver citação de PA06 abaixo). Estes *feedbacks* proporcionaram ações corretivas junto as funcionalidades com defeitos. Alguns participantes relataram como erros situações que, na verdade, indicam uma incompreensão sobre a funcionalidade ou de como operá-la. Por exemplo, um participante indicou que Porifera erra porque não está funcionando incluir foto (veja o comentário de PA01 abaixo). Na realidade, a inclusão da foto do pesquisador é feita de forma automática quando um pesquisador usa no cadastro um endereço de e-mail que possui uma foto associada ao mesmo no serviço Gravatar¹³.

PA04a *“Teve um erro na habilitação do botão Create Account. Fiz o seguinte processo: (...), mas mesmo assim não habilitou o botão, tive que dar F5 na página e preencher tudo novamente”*.

PA04b *“A ordenação só funciona uma vez. Estava pela ordem de numeração, troquei para título, quis voltar para numeração não volta... fiz outro teste com ordenação de ano, mesmo caso, não troca a ordenação”*.

¹³<http://pt.gravatar.com/>

PA04c “[As palavras com] acentos ~[til] foram importadas [com] caracteres especiais”.

PA03 “Estava dando alguns erros ao tentar adicionar uma regra de Exclude/Reject (no console log dava “Cannot ready property ‘title’ of undefined” (edit:952)”.

PA06 “Coloquei o filtro de inclusão I2, mas ele não filtrou, continuou aparecendo normal”.

PA01 “A inclusão de foto não está funcionando”.

5.3.2. Oportunidades de Melhorar Usabilidade

Algumas dificuldades apontadas pelos usuários foram categorizadas como Oportunidades de Melhorar a Usabilidade. Isto porque, não tratava de um erro propriamente dito, mas de um comportamento que poderia ser modificado para que a operação se tornasse mais fácil ou prevenindo erros, por exemplo. Alguns comentários sobre este aspecto foram: a interface não avisa o limite de caracteres no formulário de adição de critérios (ver citação de PA02), não aparecimento da opção de aguardando conforme descrito no tutorial, do usuário não conseguir ver as revisões de outros pesquisadores (ver comentários PA04a e PA04b abaixo), a escolha por um modelo GQM substituir o texto de objetivos já escrito (ver comentário de PA08 abaixo) e de não permitir uso de aspas duplas em comentários de revisão (ver citação de PA05 abaixo). Algumas destas oportunidades foram implementadas em seguida à avaliação, como, por exemplo, o aviso quando pede para usar uma frase-modelo GQM e já existe texto preenchido para o campo (Figura 7).

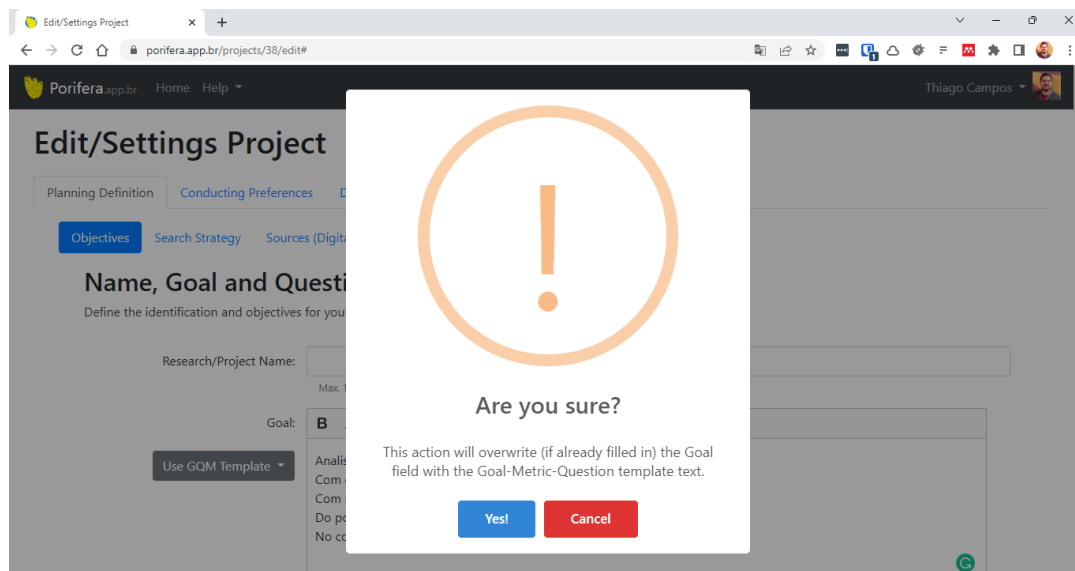


Figura 7. Aviso alertando de que já existe texto preenchido, para que o usuário confirme ou cancele a operação.

PA02 “*Não está validando o número máximo de caracteres e também não retorna mensagem avisando que ultrapassou o limite de caracteres*”.

PA04a “*Não aparece a opção de aguardando: Enquanto os demais colegas não aplicam o critério para a mesma publicação e filtro, o sistema indica que é necessário aguardar*”.

PA04b “*Estou logada em Porifera e recebi a confirmação que os 15 artigos foram finalizados: “Oi [omitido], bom dia! Finalizei os 15 artigos do primeiro filtro. Saudações cordiais, [omitido]”. Mesmo voltando na home e entrando novamente no Review não consigo ver os artigos analisados por ela. Vou sair de Porifera e fazer login novamente para verificar se aparece os artigos analisados pelo outro membro*”.

PA08 “*Eu selecionei o GQM em português, salvei. Criei as palavras-chave em português (mas queria em inglês), aí voltei para mudar o GQM para o inglês para ver se as palavras-chave também mudaria, e perdi o que já havia salvo*”.

PA05 “*Não interpreta aspas duplas*”.

5.3.3. Sugestões

Os participantes também apresentaram *feedbacks* com sugestões de melhoria ou novas funcionalidades, como: permitir editar perfil, permitir editar ou adicionar dados das publicações (ver citações PA01a e PA01b a seguir), permitir listar mais de 50 publicações por página (ver citação de PA06 abaixo), permitir copiar DOI da publicação (ver citação de PA07), limpar o formulário de convite de colaborador após o envio e escolher o idioma de strings de busca (ver citações PA08a e PA08b abaixo). Uma das sugestões, como a de copiar alguns metadados da publicação para a área de transferência, foi implementada poucos dias depois do *feedback* recebido (Figura 8).

PA01a “*Senti falta da possibilidade de edição do perfil*”.

PA01a “*Seria possível editar os dados das referências? Por exemplo a que não tem um link, inserir o link no botão open*”.

PA06 “*Acho que seria interessante poder carregar mais de 50 publicações por página, 100 estaria em uma média mais legal*”.

PA07 “*Seria interessante um botão que permitisse copiar direto o DOI da publicação, ao invés de abrir em outra página para ter que achá-lo*”.

PA08a “*Sugestão: depois de enviar o convite, limpar o form para receber as informações do próximo convidado (se houver). Mas de qualquer forma fica legal deixar ele em branco após a realização de um convite*”.

PA08b “*Agora que me toquei que se quero a string em inglês tenho que já escrever em inglês. Mas seria legal ter a opção de escolha da língua (mas imagino que deva dar um trabalhinho pq precisaria de um dicionário*”.

Publication Info ☆ ∞

#13, Collected at 03/06/2022 from ACM Digital Library
 Classified by source as: **inproceedings**
 Srinivasan, Karthik and Singh, Dineshkumar and Lonk
 Alla, Divya and Sarangi, Sanat , 2018
**Feedback System for Improving Capturing Quality
 Images Using Deep Learning Technology**
 ▶ [abstract](#)
 This seems be unique on project!

see google translate cite ↓ .bib **copy** ▼ add file link

authors
 DOI
 title
 title & authors & year
 URL

Figura 8. Botão “copy” permite ao usuário copiar algumas informações dos metadados das publicações.

5.3.4. Utilidade e Facilidade de Uso

Sobre a utilidade e facilidade de uso de Porifera, um dos participantes considerou que a ferramenta proposta é intuitiva e possui ótimas funcionalidades que ajudam no processo de RSL (ver citações PA01a e PA01b a seguir). Também disseram ser possível visualizar os dados e acessar os artigos de forma fácil (ver citações PA05a e PA05b abaixo). Além disso, um dos participantes classificou Porifera como uma solução “muito boa” (ver citação de PA08).

PA01a “A ferramenta é excelente e super intuitiva”.

PA01b “Ótimas funcionalidades. Está ajudando muito”.

PA05a “Adorei poder acessar os artigos facilmente!”.

PA05b “Fantástico poder visualizar os dados tão facilmente!”.

PA08 “(...) achei muito bom o sistema”.

5.4. Discussão sobre a avaliação

Para responder à questão de pesquisa deste estudo, escolheu-se avaliar Porifera de forma quantitativa usando o TAM e acrescentando algumas sentenças em relação ao aspecto da eficácia. Além disso, uma avaliação qualitativa utilizando alguns procedimentos de um subconjunto das etapas do método *Grounded Theory* (GT) [Corbin and Strauss 2014] foi realizada sobre dados coletados em questões abertas no mesmo questionário e em *feedbacks* fornecidos durante o uso da ferramenta. Deste modo, tem-se o uso de um método consolidado para mensurar a facilidade de uso, a utilidade e a intenção de uso de Porifera simultaneamente, em que, buscou-se uma análise qualitativa bem fundamentada, permitindo identificar problemas e encontrar possíveis soluções para eles.

A qualidade de um produto de software, segundo a norma internacional ISO/IEC 25010 [ISO 2011] e a nacional NBR 9126-110

[Associação Brasileira de Normas Técnicas 2003] é definida por um conjunto de características (atributos). Dentre estas características estão a funcionalidade, a usabilidade e a eficiência. É possível perceber que o TAM e as questões abertas propostas nesta avaliação cobrem, em parte, tais características e permitem trazer uma compreensão acerca da qualidade de Porifera também sobre o aspecto da qualidade de produto de software.

Os resultados qualitativos do questionário e, adicionalmente, do conjunto de sentenças sobre eficácia e emoção escolhida no *feedback* avaliam aspectos que podem ser comparados aos componentes da avaliação de usabilidade do método SUS, por exemplo. Entretanto, não substituí os métodos de avaliação de cobertura de requisitos, razão pela qual, estes deverão ser aplicados em trabalhos futuros. Além disso, métodos específicos para avaliação de usabilidade ou experiência de usuário podem trazer novas descobertas sobre este aspecto da qualidade e consequente aprimoramento da ferramenta.

A avaliação qualitativa realizada por meio do método GT possibilitou uma análise mais assertiva dos dados coletados e serviu tanto para referendar os números obtidos na análise quantitativa quanto para permitir enxergar falhas e pontos de melhoria na ferramenta, conforme desejado. Os achados que reportam erros e falhas na ferramenta serviram para aplicação de ações corretivas para evitar que tais problemas comprometessem o uso e a eficácia da ferramenta. As sugestões trazidas pelos usuários serviram para aprimorar alguns pontos da interface, tornando-a mais fácil ou agradável de usar.

5.4.1. Validade e Limitações

A validade desta avaliação pode estar limitada a alguns fatores. Pode haver um efeito em razão da homogeneidade dos participantes, sendo todos alunos de uma mesma instituição e disciplina - embora tenha tido alunos de diferentes níveis de ensino e de diferentes cursos.

Além disso, poucos participantes tinham experiência com outras ferramentas de apoio para RSL, o que pode ter limitado eles realizarem uma avaliação mais crítica comparando a experiência de uso de Porifera com experiências anteriores. Ainda com relação aos participantes, o fato da maioria terem experiência como iniciante em RSL poderia impactar em uma avaliação superficial da ferramenta. Para mitigar este efeito, a avaliação da ferramenta aconteceu durante o curso de disciplina que previamente introduziu os conceitos e demonstrou os passos da realização de uma RSL.

O tamanho da amostra também pode não ser ideal do ponto de vista estatístico, embora uma amostra de pelo menos 5 usuários seja considerada adequada para testes de usabilidade [Nielsen 2000, Nielsen 2012]. A análise de emoções pode apresentar imperfeições devido ao protocolo de análise de conteúdo [Santos 2012] seguido. Desta forma, aparentemente pode parecer imprecisa, porque foi baseada na interpretação do verbo usado no comentário fornecido pelo usuário (caso da análise qualitativa, seção 5.2.6) ou na classificação dada pelo emoji escolhido (no caso da análise quantitativa, seção 5.1). Para uma melhor interpretação acerca das emoções de fato expressadas pelos usuários talvez fosse necessário outros mecanismos de coleta de dados, tais como registro

em vídeo ou sinais neurofisiológicos.

6. CONCLUSÕES E TRABALHOS FUTUROS

Este artigo apresentou a importância do aspecto da colaboração em uma ferramenta de apoio para RSL e o modo com que tais ferramentas costumam ser avaliadas. Incluiu uma discussão sobre os métodos utilizados para avaliar estas ferramentas e apresentou uma avaliação usando uma combinação de um método quantitativo e qualitativo para investigar a qualidade de uma nova ferramenta de apoio a RSL, bem como para identificar problemas e pontos de melhoria.

Nesta avaliação de Porifera, foi escolhido o TAM para obter diferentes percepções sobre a aceitação da ferramenta. Além disso, uma análise qualitativa foi realizada por meio do método GT para permitir enxergar falhas e pontos de melhoria. Os resultados da avaliação indicaram a utilidade percebida e a facilidade de uso da Porifera, além da eficácia dos recursos para apoiar a atividade colaborativa de uma RSL. A análise qualitativa confirmou que a ferramenta é intuitiva, prática e fácil de usar e que, usá-la traz agilidade e facilidades a realização de uma RSL. Foi percebido que seu uso contribuiu com o aprendizado e a execução de uma RSL colaborativa(o), levando os participantes a terem intenção de continuar usando-a e até mesmo recomendando-a a outros pesquisadores. Ressalta-se que o estudo possui limitações, como o fato de poucos participantes terem experiência prévia na realização de RSL e uma análise de emoções baseada em conteúdo verbal.

Este trabalho trouxe, portanto, algumas contribuições, como: uma análise de sistemas de informações voltados a apoiar RSL que procuram propor soluções para realização de uma pesquisa colaborativa; um levantamento dos métodos de avaliação de qualidade que podem ser usados para avaliar este tipo de sistemas; uma classificação destes métodos de avaliação; uma breve apresentação de uma nova ferramenta de apoio à RSL; uma avaliação de qualidade combinando métodos quantitativos e qualitativos para um SI colaborativo, cobrindo aspectos de utilidade, facilidade, eficácia, intenção de uso e usabilidade em geral e; uma avaliação e análise com olhar voltado nas emoções e experiência do usuário.

Os próximos passos incluem a implementação de alguns recursos planejados para a ferramenta, como critérios de exclusão de decisão automatizada com base em metadados (ano, idioma, natureza da publicação, por exemplo); formulários ou integração com serviço externo para etapa de extração de dados; relatórios de resultados da seleção de estudos e; exportação de publicações aceitas para uso em gerenciadores de referências bibliográficas. Também pode ser fruto de trabalho futuro a proposta de um método de avaliação de ferramentas de apoio a RSL colaborativa.

Novas avaliações deverão ocorrer, com uma amostra maior de usuários e por um tempo maior de execução de RSL. Também espera-se avaliação por especialistas, objetivando verificar a cobertura de requisitos para ferramentas de RSL, conforme proposto por Marshall et al. [2014] e Al-Zubidy et al. [2019] e compará-la com outras soluções do gênero. Também pretende-se realizar avaliações de usabilidade, acessibilidade e experiência do usuário (UX) e implementar diferentes níveis de teste de código.

Referências

- Al-Zubidy, A. and Carver, J. C. (2019). Identification and prioritization of SLR search tool requirements: an SLR and a survey. *Empirical Software Engineering*, 24(1):139–169.
- Al-Zubidy, A., Carver, J. C., Hale, D. P., and Hassler, E. E. (2017). Vision for SLR tooling infrastructure: Prioritizing value-added requirements. *Information and Software Technology*, 91:72–81.
- Associação Brasileira de Normas Técnicas (2003). NBR ISO/IEC 9126-1: Engenharia de software - Qualidade de produto - Parte 1: Modelo de qualidade.
- Bangor, A., Kortum, P., and Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating ScaleJUS. *Journal of Usability Studies*, 4(3):114–123.
- Barn, B. S., Raimondi, F., Athappian, L., and Clark, T. (2014). Slrtool: A tool to support collaborative systematic literature reviews. In *ICEIS 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems*, volume 2, pages 440–447, Lisbon, Portugal. SciTePress.
- Basili, V. R. and Caldiera, G. (2000). The Goal Question Metric Paradigm. *Encyclopedia of Software Engineering - 2 Volume Set*, 2:528–532.
- Bigendako, B. and Syriani, E. (2018). Modeling a tool for conducting systematic reviews iteratively. In *Proceedings of the 6th International Conference on Model-Driven Engineering and Software Development - MODELSWARD*, volume 1, pages 552–559, Madeira, Portugal. SciTePress.
- Bowes, D., Hall, T., and Beecham, S. (2012). Slurp: A tool to help large complex systematic literature reviews deliver valid and rigorous results. In *Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies*, EAST '12, page 33–36, New York, NY, USA. Association for Computing Machinery.
- Brooke, J. (1996). SUS: A 'Quick and Dirty' Usability Guide. In *Usability Evaluation In Industry*, chapter 21, page 6. CRC Press, London, UK, 1st edition.
- Bustos Navarrete, C., Morales Malverde, M. G., Salcedo Lagos, P., and Díaz Mujica, A. (2018). Buhos: A web-based systematic literature review management software. *SoftwareX*, 7:360–372.
- Campos, T. P. d., Damasceno, E. F., and Valentim, N. M. C. (2022). Proposal and evaluation of a collaborative is to support systematic reviews and mapping studies. In *XVIII Brazilian Symposium on Information Systems*, SBSI, New York, NY, USA. Association for Computing Machinery.
- Corbin, J. and Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publishing, San Jose, USA, 4th edition.
- Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3):475–487.

- de Brito, J. V. d. C. S. and Ramos, A. S. M. (2019). Limitações dos Modelos de Aceitação da Tecnologia: um Ensaio Sob uma Perspectiva Crítica. *Gestão.Org - Revista Eletrônica de Gestão Organizacional*, 17(0):210–220.
- Fan, M., Shi, S., and Truong, K. N. (2020). Practices and challenges of using think-aloud protocols in industry: An international survey. *Journal of Usability Studies*, 15(2):85–102.
- Fernández-Sáez, A., Bocco, M., and Romero, F. (2010). SLR-TOOL - A Tool for Performing Systematic Literature Reviews:. In *Proceedings of the 5th International Conference on Software and Data Technologies*, pages 157–166, University of Piraeus, Greece. SciTePress - Science and and Technology Publications.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.
- Hassler, E., Carver, J. C., Hale, D., and Al-Zubidy, A. (2016). Identification of SLR tool needs - Results of a community workshop. *Information and Software Technology*, 70:122–129.
- Hernandes, E., Zamboni, A., Fabbri, S., and Di Thommazo, A. (2012). Using GQM and TAM to evaluate StArt – a tool that supports Systematic Review. *CLEI Electronic Journal*, 15(1):3–3.
- Hinderks, A., Jose, F., Mayo, D., Thomaschewski, J., and Escalona, M. J. (2020). An SLR-Tool: Search Process in Practice : To conduct and manage Systematic Literature Review (SLR). In *Proceedings - 2020 ACM/IEEE 42nd International Conference on Software Engineering: Companion, ICSE-Companion 2020*, pages 81–84, Seoul, South Korea. Institute of Electrical and Electronics Engineers Inc.
- ISO (2011). ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. Technical report, International Organization for Standardization, Geneva.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical report, Keele University and University of Durham.
- Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O. P., Turner, M., Niazi, M., and Linkman, S. (2010). Systematic literature reviews in software engineering-A tertiary study.
- Kitchenham, B. A., Budgen, D., and Brereton, P. (2015). *Evidence-Based Software Engineering and Systematic Reviews*. Chapman and CRC Press, Boca Raton, USA, 1st edition.
- Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., and Wilhelm, R. (2018). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, 7(1):1–17.
- Marchezan, L., Bolfe, G., Rodrigues, E., Bernardino, M., and Basso, F. P. (2019). Thoth: A Web-based Tool to Support Systematic Reviews. In *ACM/IEEE International Sym-*

- posium on Empirical Software Engineering and Measurement (ESEM)*, volume 1, pages 3–8, Porto de Galinhas, Brazil. Institute of Electrical and Electronics Engineers (IEEE).
- Marshall, C. (2016). *Tool support for systematic reviews in software engineering*. PhD thesis, Keele University, Keele, Newcastle, UK.
- Marshall, C. and Brereton, P. (2015). Systematic review toolbox: A catalogue of tools to support systematic reviews. In *EASE '15: Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, volume 1, pages 1–6, Nanjing, China. Association for Computing Machinery.
- Marshall, C., Brereton, P., and Kitchenham, B. (2014). Tools to support systematic reviews in software engineering: A feature analysis. In *EASE '14: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, pages 1–10, London, UK. Association for Computing Machinery.
- Molléri, J. S. and Benitti, F. B. V. (2015). SESRA - A web-based automated tool to support the systematic literature review process. In *EASE '15: Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*, volume 1, pages 1–6, Nanjing, China. Association for Computing Machinery.
- Nielsen, J. (2000). Why You Only Need to Test with 5 Users. <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>. Accessed: 2022-09-15.
- Nielsen, J. (2012). How Many Test Users in a Usability Study? <https://www.nngroup.com/articles/how-many-test-users/>. Accessed: 2022-09-15.
- Okoli, C. and Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *SSRN Electronic Journal*.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1):1–10.
- Parsifal Ltd. (2018). About Parsifal. <http://parsif.al/about/>. Accessed: 2021-07-12.
- Petersen, K. and Ali, N. B. (2011). Identifying strategies for study selection in systematic reviews and maps. *International Symposium on Empirical Software Engineering and Measurement*, pages 351–354.
- Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *Electronic Workshops in Computing*. BCS Learning & Development.
- Petticrew, M. and Roberts, H. (2005). *Systematic Reviews in the Social Sciences: A Practical Guide*. Wiley-Blackwell, 1st edition.
- Santos, F. M. d. (2012). Análise de conteúdo: a visão de laurence bardin. *Revista Eletrônica de Educação*, 6(1):383–387.

- Silva, R. d. S. e. and Paes, Â. T. (2012). Por Dentro da Estatística: teste de concordância de Kappa. *Educ. Contin. Saúde einstein*, 10(4):165–166.
- Stefanovic, D., Havzi, S., Nikolic, D., Dakic, D., and Lolic, T. (2021). Analysis of the Tools to Support Systematic Literature Review in Software Engineering. *IOP Conference Series: Materials Science and Engineering*, 1163(1):012013.
- Sutton, A., Clowes, M., Preston, L., and Booth, A. (2019). Meeting the review family: exploring review types and associated information retrieval requirements. *Health Information Libraries Journal*, 36(3):202–222.