

Revisão sistemática da literatura em engenharia de software

Teoria e prática

Revisão sistemática da literatura em engenharia de software

Teoria e prática

Katia Romero Felizardo

Elisa Yumi Nakagawa

Sandra Camargo Pinto Ferraz Fabbri

Fabiano Cutigi Ferrari

ELSEVIER

© 2017, Elsevier Editora Ltda.

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998.

Nenhuma parte deste livro, sem autorização prévia por escrito da editora, poderá ser reproduzida ou transmitida sejam quais forem os meios empregados: eletrônicos, mecânicos, fotográficos, gravação ou quaisquer outros.

ISBN (versão digital): 978-85-352-8597-0

Copidesque: Tássia Hallais

Elsevier Editora Ltda.

Conhecimento sem Fronteiras

Rua Sete de Setembro, 111 – 16º andar
20050-006 – Centro – Rio de Janeiro – RJ
Rua Quintana, 753 – 8º andar
04569-011 – Brooklin – São Paulo – SP

Serviço de Atendimento ao Cliente
0800 026 53 40
atendimento1@elsevier.com

Consulte nosso catálogo completo, os últimos lançamentos e os serviços exclusivos no site www.elsevier.com.br

NOTA

Muito zelo e técnica foram empregados na edição desta obra. No entanto, podem ocorrer erros de digitação, impressão ou dúvida conceitual. Em qualquer das hipóteses, solicitamos a comunicação ao nosso serviço de Atendimento ao Cliente para que possamos esclarecer ou encaminhar a questão.

Para todos os efeitos legais, nem a editora, nem os autores, nem os editores, nem os tradutores, nem os revisores ou colaboradores, assumem qualquer responsabilidade por qualquer efeito danoso e/ou malefício a pessoas ou propriedades envolvendo responsabilidade, negligência etc. de produtos, ou advindos de qualquer uso ou emprego de quaisquer métodos, produtos, instruções ou ideias contidos no material aqui publicado.

A Editora



Colaboradores

Bento Rafael Siqueira

Departamento de Computação

Universidade Federal de São Carlos – UFSCar

bento.siqueira@dc.ufscar.br

Brauner Roberto do Nascimento Oliveira

Departamento de Sistemas de Computação

Instituto de Ciências Matemáticas e de Computação – ICMC

Universidade de São Paulo – USP

brauner@usp.br

Cleiton Rodrigo Queiroz Silva

Departamento de Computação

Universidade Federal de São Carlos – UFSCar

cleiton.silva@dc.ufscar.br

Elis Cristina Montoro Hernandes

Departamento de Computação

Universidade Federal de São Carlos – UFSCar

elis_hernandes@dc.ufscar.br

Elisa Yumi Nakagawa

Departamento de Sistemas de Computação

Instituto de Ciências Matemáticas e de Computação – ICMC

Universidade de São Paulo – USP

elisa@icmc.usp.br

Érica Ferreira de Souza

Departamento de Computação

Universidade Tecnológica Federal do Paraná – UTFPR/Cornélio Procópio

ericasouza@utfpr.edu.br

Fábio Roberto Octaviano

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP/São Carlos

foctaviano@ifsp.edu.br

Fabiano Cutigi Ferrari

Departamento de Computação

Universidade Federal de São Carlos – UFSCar

fabiano@dc.ufscar.br

Katia Romero Felizardo

Departamento de Computação

Universidade Tecnológica Federal do Paraná – UTFPR/Cornélio Procópio

katiascannavino@utfpr.edu.br

Lucas Bueno Ruas de Oliveira

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP/São Carlos

lucas.oliveira@ifsp.edu.br

Milena Guessi

Departamento de Sistemas de Computação

Instituto de Ciências Matemáticas e de Computação – ICMC

Universidade de São Paulo – USP

milena@icmc.usp.br

Ricardo de Almeida Falbo

Departamento de Informática

Universidade Federal do Espírito Santo – UFES

falbo@inf.ufes.br

Sandra Camargo Pinto Ferraz Fabbri

Departamento de Computação

Universidade Federal de São Carlos – UFSCar

sfabbri@dc.ufscar.br

Apresentação

A Revisão Sistemática (RS) tem sido largamente explorada nos últimos anos na área de Engenharia de Software (ES), trazendo consideráveis contribuições para os avanços das pesquisas e práticas nessa área. Este livro tem, então, como principal objetivo, ensinar RS por meio de perguntas, respostas e exemplos, e é fruto das perguntas coletadas em nossas palestras, cursos e também durante a condução de diversas RSs nos últimos anos, representando as dúvidas mais comuns sobre RS. O livro é prático e didático, e poderá ser utilizado como guia para a melhor compreensão do passo a passo e dos detalhes que envolvem o processo de condução da RS. Esperamos que atenda a demanda por um livro de consulta rápida e objetiva e que venha a ser de utilidade para pesquisadores e alunos de pós-graduação que precisem conduzir suas RSs como parte de seus projetos de pesquisa. Também idealizamos um livro que poderá ser adotado por professores como um material de apoio em suas aulas, tais como nas de metodologia de pesquisa.

Esta obra pode ser lida de duas maneiras: (1) leitura sequencial, do primeiro ao último capítulo e, ao final, o leitor terá informações que poderão auxiliá-lo na condução de sua própria revisão; e (2) leitura de perguntas e respostas pontuais, em uma ordem definida pelos interesses do próprio leitor. As características particulares deste livro, que o diferenciam de outros publicados na mesma linha, incluem:

- Organização em capítulos com objetivos claramente estabelecidos e que podem ser lidos de forma sequencial ou na ordem de preferência do leitor.
- Capítulos contendo texto discursivo, imagens, tabelas e também referências a outros trabalhos diretamente relacionados.
- Explicação clara dos aspectos práticos e teóricos do processo de RS.
- Orientação a pesquisadores, incluindo alunos de pós-graduação, que já conduziram ou estão iniciando-se na condução de RS, inclusive em outras áreas da computação além da ES.

Cada capítulo inclui considerações iniciais, nas quais se apresentam o contexto e os objetivos de aprendizagem. O Capítulo 1 é dedicado aos principais conceitos sobre RS, dando ênfase ao processo para a condução de revisões. No Capítulo 2, é detalhado o protocolo da RS utilizado tanto para formalizar a revisão quanto para acompanhar e documentar sua execução. No Capítulo 3 discutimos em detalhes o quão abrangente e completa deve ser a busca por estudos a serem incluídos em uma revisão. O Capítulo 4 aborda o estabelecimento dos critérios para a seleção dos estudos identificados a partir da busca. O Capítulo 5 discute os métodos que

podem ser utilizados para analisar e sumarizar os resultados de uma RS. No Capítulo 6 apresenta-se brevemente o processo de Mapeamento Sistemático (MS), visando compará-lo ao processo de RS, o foco deste livro. E o Capítulo 7 destaca as considerações finais, enfatizando os desafios associados à realização de RSs.

Por fim, vale a pena ressaltar que este livro se tornou possível por meio da colaboração de pesquisadores de diferentes universidades brasileiras, incluindo a Universidade de São Paulo (ICMC-USP – São Carlos/SP), a Universidade Federal de São Carlos (UFSCar – São Carlos/SP), o Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP – São Carlos/SP), a Universidade Tecnológica Federal do Paraná (UTFPR – Cornélio Procópio/PR) e a Universidade Federal do Espírito Santo (UFES – Vitória/ES).

Gostaríamos de agradecer a todos os colaboradores dos capítulos deste livro e ao Professor Doutor Guilherme Horta Travassos, da COPPE/UFRJ, pioneiro na área de Engenharia de Software Experimental no Brasil e um dos mais renomados pesquisadores nessa área, por prontamente aceitar o nosso convite para escrever o prefácio. Gostaríamos de estender nossos mais sinceros agradecimentos ao Professor Doutor José Carlos Maldonado, do ICMC/USP, que foi o responsável por introduzir a RS em nossas pesquisas, idealizou a escrita deste livro e, posteriormente, nos incentivou a escrevê-lo.

Desejamos a todos uma boa leitura!

Katia Romero Felizardo

Elisa Yumi Nakagawa

Sandra Camargo Pinto Ferraz Fabbri

Fabiano Cutigi Ferrari

Prefácio

A busca pela evidência tem sido um desafio constante em Engenharia de Software. Ao longo dos últimos 30 anos, diferentes estratégias de observação têm sido aplicadas para apoiar a investigação de nosso objeto de estudo mais importante, o software. Bem verdade que ainda não conseguimos responder a várias perguntas básicas sobre ele, nem mesmo conseguimos responder de forma precisa o que ele é. Entretanto, temos evoluído nossas habilidades científicas e demonstrado nossa capacidade de investigação através da realização de vários estudos primários associados a diferentes problemas e fenômenos do software. Quanto resultado disponível! Quanta indecisão ainda há no campo! Quanta dificuldade para sua compreensão! Como descobrir o que existe? Como agrregar esses resultados?

Estudos secundários intencionam apoiar a observação e agregação dos resultados dos estudos primários. A existência de diferentes resultados relacionados com problemas “em comum” justificam a elaboração de estudos desta natureza. Não tem muito tempo que iniciamos a execução de estudos secundários em Engenharia de Software. Por volta de 2003, Barbara Kithchenham, Emilia Mendes e eu iniciamos as discussões dos primeiros protocolos de revisão sistemática da literatura aplicados à Engenharia de Software. Outros pesquisadores tais como Magne Jørgensen e Tøre Dyba também iniciavam discussões semelhantes. Estávamos bastante motivados pelos resultados que vinham sendo obtidos em outras áreas de ciência, tais como medicina, e isso nos levava a trabalhar no sentido de buscar internalizar os conceitos deste tipo estudo em Engenharia de Software. Quantos desafios! A falta de uma taxonomia consistente, uma agenda de pesquisa bem-definida (parece que pouco mudou desde então...) e sistemas de buscas robustos faziam as revisões sistemáticas difíceis e de longa duração. Assim, a experiência em preparar e executar diferentes revisões sistemáticas aliadas às dificuldades naturais da área nos levou a buscar níveis intermediários de trabalho. Em algum momento do tempo, estruturas mais leves (tais como mapeamentos sistemáticos ou *quasi* revisões sistemáticas da literatura) passaram a ser utilizadas e a contribuir com a organização do conhecimento em Engenharia de Software. A realização desses estudos nos permitiu destacar a falta de evidência relacionada com as diferentes tecnologias de software! Quanto a desenvolver e evoluir!

Muito conhecimento metodológico relacionado com estudos secundários tem sido gerado, com o estabelecimento de práticas e novas estratégias de apoio. Percebe-se uma evolução natural dos sistemas de busca, mesmo que ainda sejamos deficitários em discussões relacionadas com a taxonomia em Engenharia de Software e não consigamos definir uma

agenda de pesquisa comum que nos permita investir esforços combinados e com foco experimental bem definido. Entretanto, muito se aprendeu e se aprende. O engajamento dos pesquisadores em Engenharia de Software neste movimento foi de fundamental importância. Vários materiais foram produzidos ao longo deste período e apresentados como relatórios técnicos, artigos científicos, modelos ontológicos, algoritmos de busca, dentre outros. Porém, a diversidade deste material e suas mais diferentes estruturas de apresentação acabam dificultando o acesso aos conceitos básicos por aqueles interessados no aprendizado de revisões sistemáticas da literatura e suas derivações em Engenharia de Software.

O livro *Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática*, organizado pelos pesquisadores Katia Romero Felizardo, Elisa Yumi Nakagawa, Sandra Camargo Pinto Ferraz Fabbri e Fabiano Cutigi Ferrari, representa um diferencial interessante nessa área. Ao reunir os conceitos básicos deste tipo de estudo de uma forma simples, rigorosa e didática, ele contribui para a disseminação deste conhecimento entre os pesquisadores de Engenharia de Software interessados nessa estratégia de investigação. A estrutura em capítulos autocontidos facilita a leitura e o aprendizado, e permite delinear corretamente os conceitos envolvidos com as revisões sistemáticas da literatura e sua utilização em prol do avanço da área.

Muito temos ainda a aprender e evoluir em Engenharia de Software. Porém, o conhecimento adquirido até então pode e deve ser corretamente aplicado no sentido de alavancar nossa evolução. Por isso, materiais como este são importantes e trazem benefícios intangíveis para a nossa comunidade. Espero que a leitura seja proveitosa e o material apoie a elaboração de seus estudos secundários! Parabéns aos organizadores e colaboradores! Boa Revisão Sistemática da Literatura!

Guilherme Horta Travassos

Pesquisador CNPq

Professor Titular de Engenharia de Software
Programa de Engenharia de Sistemas e Computação
COPPE/UFRJ

Sumário

1.	Revisão sistemática.....	14
2.	Protocolo da revisão sistemática.....	24
3.	Identificação de estudos.....	39
4.	Seleção e avaliação de estudos	57
5.	Síntese dos dados e apresentação dos resultados	75
6.	Mapeamento Sistemático.....	92
7.	Considerações finais	112
8.	Referências Bibliográficas.....	115

Lista de Figuras

Figura 1.1: Fases e atividades do processo de RS.....	17
Figura 3.1: Representação em conjunto da estratégia Pearl Growing.	42
Figura 4.1: Etapas da seleção de estudos.	63
Figura 4.2: Etapas do cálculo do teste Kappa.	70
Figura 4.3: Sugestão para a documentação da seleção de estudos.	72
Figura 6.1: Relação entre tipos de pesquisa e métodos de pesquisa (adaptado de [49]).	105

Lista de Tabelas

Tabela 1.1 Ferramentas de apoio à RS [24] [25] [26].	20
Tabela 2.1: Seções e itens que compõem o protocolo.....	25
Tabela 2.2: Critérios PICO referentes a um exemplo da área médica.	28
Tabela 2.3: Critérios PICO na área de ES - descrição de processos.....	29
Tabela 2.4: Critérios PICO na área de ES - orientação a serviço e sistemas robóticos.	31
Tabela 3.1: Principais bases bibliográficas em Ciência da Computação.	46
Tabela 4.1: Resultado da seleção contendo estudos incluídos e excluídos.....	69
Tabela 5.1: Comparação entre métodos de síntese.	83
Tabela 6.1: Diferenças entre entre MSs e RSs.	94
Tabela 6.2: Esquema de classificação de pesquisa. Fonte: [79].	104

1. Revisão sistemática

Érica Ferreira de Souza
Katia Romero Felizardo
Ricardo de Almeida Falbo

A Revisão Sistemática (RS) tem como objetivo identificar, analisar e interpretar as evidências disponíveis relacionadas a um particular tópico de pesquisa ou fenômeno de interesse. O objetivo deste capítulo é apresentar uma visão abrangente sobre RS e ressaltar as principais diferenças entre uma RS e uma revisão tradicional da literatura. O processo de RS é detalhado, assim como suas fases, atividades e os recursos necessários para a sua execução. Serão discutidas também sobre a confiabilidade, vantagens, desvantagens e limitações das RSs.

Questão 1: O que é uma RS?

Em 2004, Kitchenham, Dybå e Jørgensen [1] propuserem a Engenharia de Software Baseada em Evidência (ESBE) como um meio de aperfeiçoar a tomada de decisões relativas ao desenvolvimento e manutenção de software, por meio da integração de evidências de pesquisa com experiências práticas e valores humanos. O objetivo principal da ESBE é diminuir a lacuna existente entre a pesquisa e a prática na Engenharia de Software (ES) e, para tal, disponibiliza meios para construir um corpo de conhecimento sobre quando, como e quais processos, tecnologias, ferramentas, entre outros, são adequados para serem utilizados na ES [1] [2] [3]. Nesse contexto, a Revisão Sistemática da Literatura (RSL), ou simplesmente Revisão Sistemática (RS), é um dos principais meios para sumarizar evidências de pesquisa [4]. A RS tem se tornado cada vez mais popular, conforme pode ser constatado em trabalhos como [4] [5].

A RS tem como objetivo identificar, selecionar, avaliar, interpretar e sumarizar estudos disponíveis considerados relevantes para um tópico de pesquisa ou fenômeno de interesse [1] [6]. Os estudos individuais que contribuem para uma RS são chamados de estudos primários, enquanto a RS em si é considerada um estudo secundário [7]. Na mesma linha que as RSs, os Mapeamentos Sistemáticos (MSs) também são estudos secundários e, nesse caso, eles têm como objetivo identificar e classificar os estudos primários relacionados a um tópico de pesquisa, sendo então uma revisão mais ampla das evidências disponíveis de um dado tópico. No Capítulo 6, são apresentados mais detalhes sobre o MS.

De um modo geral, na área de ES, os estudos primários visam propor novas soluções e tecnologias (incluindo-se, por exemplo, os processos, métodos, técnicas e abordagens), e/ou então caracterizar uma determinada solução ou tecnologia em uso dentro de um contexto

específico. Nesse último caso, encontram-se os estudos primários que apresentam experimentos, estudos de caso e *surveys* [8]. Por outro lado, um estudo secundário analisa os estudos primários com o objetivo de sintetizar as evidências relacionadas a um tópico de pesquisa. De acordo com Brereton et al. [9], a sumarização dos resultados dos estudos primários por meio de estudos secundários é um mecanismo de pesquisa valioso por prover conhecimento de um dado tópico e apoiar na identificação de temas para pesquisa futura.

A RS é conduzida por meio de um processo composto por uma sequência de fases bem definidas. Para maiores detalhes sobre o processo de RS, leia a Questão 3 deste capítulo. A RS apresenta uma avaliação considerada justa sobre um tópico de pesquisa, uma vez que utiliza uma forma de revisão rigorosa, confiável e passível de auditoria [10]. Além disso, uma RS baseia-se em um protocolo previamente definido [6], ou seja, um documento que formaliza a execução da RS desde a definição das questões de pesquisa passando pela estratégia de busca até o relatório final. O protocolo tem como objetivo reduzir a possibilidade da ocorrência de vieses [10]. As fases do processo de RS devem ser explicitamente definidas no protocolo, assim como os critérios e estratégias de seleção, extração e sumarização dos dados, para que outros pesquisadores possam reproduzir a RS, adotando os mesmos critérios e estratégias [6]. Para maiores detalhes sobre o protocolo da RS, leia o Capítulo 2.

Por outro lado, a realização de uma revisão de forma não sistemática (ou seja, revisão informal da literatura ou revisão tradicional da literatura) está sujeita a vieses de investigação, tais como na formulação das questões de pesquisa, coleta, avaliação, interpretação e apresentação dos dados, que pode ser influenciada por interesses pessoais dos pesquisadores envolvidos, levando, por fim, a resultados pouco confiáveis [6] [11]. As revisões informais da literatura, que são conduzidas sem um planejamento estabelecido a priori, caracterizam-se por serem pouco abrangentes, não passíveis de repetição, pouco confiáveis e com a qualidade dependente da experiência dos pesquisadores [11].

Uma das razões pelas quais RSs têm sido conduzidas em ES é devido às suas vantagens quando comparadas às revisões informais, o que inclui a redução de vieses nos resultados e a possibilidade de identificar e combinar as principais diferenças entre os dados dos vários estudos selecionados na revisão. Outra vantagem é a identificação de lacunas na pesquisa atual, o que pode sugerir novos temas de investigação e fornecer uma direção para posicionar adequadamente esses temas no contexto da pesquisa existente [6]. Outros benefícios da RS são apresentados na Questão 6 deste capítulo.

Questão 2: Qual a origem da RS?

A RS teve origem na área da Medicina, na qual é considerada como um método de pesquisa chave para apoiar a pesquisa baseada em evidências [12]. A pesquisa baseada em evidências indicou que a opinião de especialistas baseada apenas na sua experiência médica não era tão confiável quanto com base em resultados de experimentos científicos, ou seja, com base nas evidências [1] [4]. Os resultados positivos do uso de RSs fizeram com que muitos especialistas passassem a adotar esse método de pesquisa. Além disso, desde 1992, observa-se que o número de artigos e revistas com interesse em RSs na área médica tem também crescido [12] [13]. O sucesso da pesquisa baseada em evidências aplicada na Medicina fez com que outras áreas de pesquisa adotassem essa abordagem, incluindo, por exemplo, a Economia, a Criminologia, a Política Social, a Enfermagem e, recentemente, a ES [9] [14].

Em 2004, por meio de uma analogia com o que é praticado na área da Medicina, Kitchenham et al. [10] propuseram diretrizes e um processo para realizar RSs na ES. De acordo com Zhang e Babar [14], a RS tornou-se um método de pesquisa bem conhecido para sumarização de evidências na ES e também na Ciência da Computação. Dessa forma, pode-se observar que o número de RSs publicadas tem crescido de forma significativa nos últimos anos [4] [5] [14] [15].

Questão 3: Como conduzir uma RS e quais são as diretrizes existentes?

Uma RS é conduzida por meio de um processo que envolve três fases [6] [9] [10] [11]: *Planejamento, Condução e Publicação dos resultados*. Essas fases, bem como as respectivas atividades, apresentadas na Figura 1.1 **Figura 1.1**, podem ser conduzidas de modo iterativo. A seguir, as três fases são apresentadas em mais detalhes.

A fase de planejamento tem como objetivo identificar a real necessidade, ou seja, a motivação para a execução de uma RS. No entanto, antes de iniciar o planejamento da revisão, é fundamental identificar se já existem estudos secundários no mesmo tema. Essa identificação pode ser feita por meio de uma Revisão Terciária, ou seja, uma revisão cujos estudos considerados são estudos secundários, ou seja, outras RSs [7]. Revisões terciárias podem ser executadas utilizando o mesmo processo adotado para uma RS. Bons exemplos de revisões terciárias na área de ES podem ser encontrados em [15] [16] [17]. Caso não haja um estudo secundário sobre o tema e esse tema seja de relevância para a comunidade científica da área, justifica-se a realização da RS.

Logo após identificar a necessidade da realização de uma RS, é definido o protocolo da revisão que é o elemento essencial para executar a revisão. O protocolo visa minimizar vieses que possam vir a ser cometidos pelo pesquisador [7] [9]. O protocolo especifica as questões de pesquisa, a estratégia que será utilizada para conduzir a RS, os critérios para a seleção dos estudos, e como os dados serão extraídos dos estudos e, em seguida, como os dados serão sintetizados.

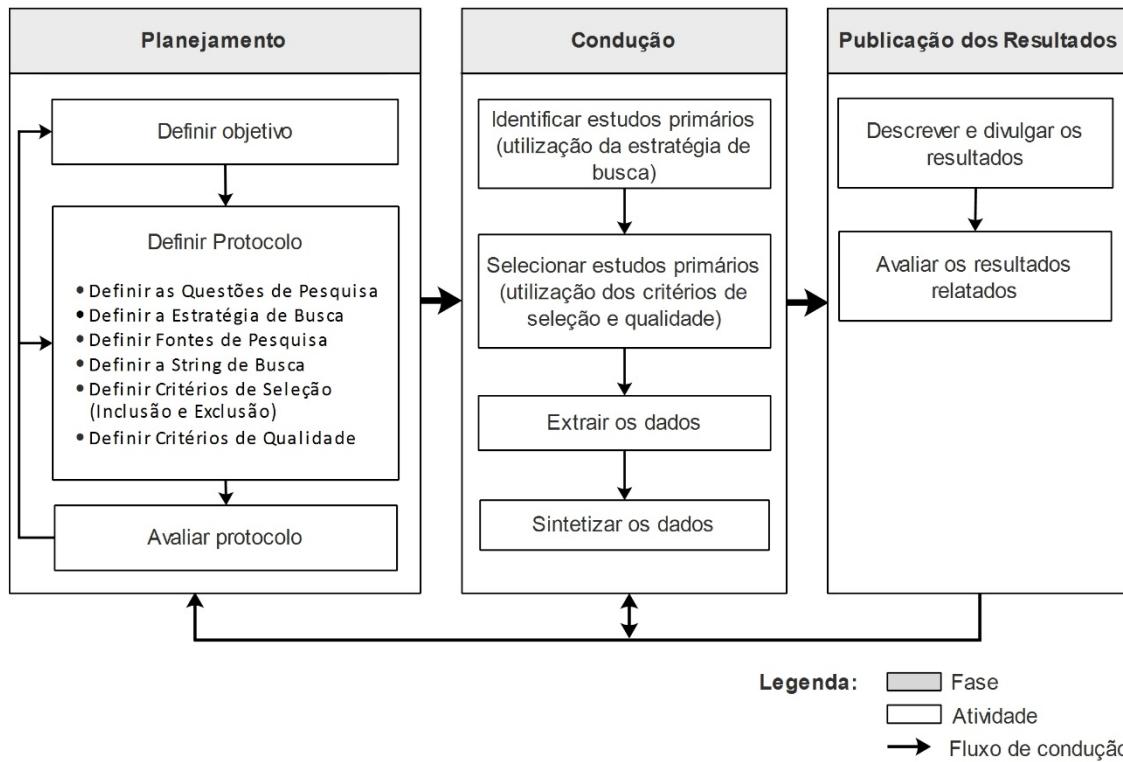


Figura 1.1: Fases e atividades do processo de RS.

Vale destacar que a qualidade do protocolo impacta diretamente a qualidade da RS. Dessa forma, o protocolo deve ser avaliado antes que se prossiga com a revisão [7]. Essa avaliação deve ser realizada por meio do teste do protocolo, também chamado de teste piloto, cujo objetivo é verificar a viabilidade de execução da revisão, permitindo também, com base nos resultados do teste, identificar modificações que sejam necessárias. É essencial que para o sucesso do teste piloto se tenha disponível um grupo de controle, ou seja, um conjunto de estudos primários que devem ser retornados a partir da condução da revisão. O grupo de controle pode ser criado por meio de uma revisão informal, realizada antes da RS, e/ou por meio da sugestão de especialistas da área. Detalhes sobre a criação e avaliação do protocolo de RS são apresentados no Capítulo 2.

Em função da natureza iterativa do processo de RS [18], os itens que compõem o protocolo podem ser refinados. Por exemplo, como é possível observar na Figura 1.1, durante a seleção

dos estudos, novos termos não inicialmente considerados na *string* de busca podem ser identificados, levando ao refinamento da *string* de busca e, posteriormente, uma nova execução das buscas por estudos primários.

Após a avaliação do protocolo, inicia-se a fase de condução da RS. Durante essa fase, os estudos são identificados utilizando-se uma estratégia de busca ampla, ou seja, que seja capaz de identificar todos os estudos primários disponíveis e relacionados ao tema de pesquisa em questão. Detalhes sobre a busca por estudos são apresentados no Capítulo 3.

Uma vez identificados, os estudos precisam ser selecionados por meio da aplicação de critérios de seleção (que são os critérios de inclusão e de exclusão), e podem ser avaliados pelos critérios de qualidade. Os critérios de seleção devem especificar as principais características e/ou conteúdos que os estudos devem ter para serem incluídos ou excluídos. Já os critérios de qualidade têm como objetivo avaliar aspectos metodológicos dos estudos, ou seja, podem ser avaliados aspectos como a relevância do tema de pesquisa e o uso de métodos que conduzam aos objetivos propostos no estudo. Os critérios de seleção e de qualidade devem estar definidos no protocolo da revisão; porém, podem ser refinados durante a atividade de seleção dos estudos [7]. No Capítulo 4 são apresentados detalhes da atividade de seleção dos estudos.

Após a atividade de seleção, os dados contidos nos estudos incluídos devem ser extraídos e sintetizados. Formulários de extração de dados são utilizados para coletar os dados que sejam necessários para responder às questões de pesquisa da revisão e para facilitar posteriormente as análises e síntese dos resultados.

A última fase do processo de RS está relacionada à escrita dos resultados da RS, que devem ser divulgados aos potenciais interessados [7]. Esses resultados podem ser divulgados por meio de relatórios técnicos, artigos de revistas ou conferências, em capítulos de livros ou como uma seção de um trabalho de conclusão de curso (tipicamente uma dissertação de mestrado ou uma tese de doutorado). Os resultados apresentados em artigos de periódicos e conferências, bem como em trabalhos de conclusão de curso passam geralmente por uma avaliação de revisores considerados especialistas no tópico de pesquisa. Porém, os relatórios técnicos geralmente não são submetidos a uma avaliação independente. Detalhes sobre como sintetizar e apresentar os resultados de uma RS são apresentados no Capítulo 5.

Questão 4: Quais são os recursos necessários para executar uma RS?

A execução de uma RS requer um conjunto de recursos, mais especificamente, recursos humanos, de software e materiais. Uma breve descrição de cada um desses recursos é apresentada a seguir.

Para a execução de uma RS, é necessário definir quantas e quais pessoas irão compor a equipe da revisão e o papel que cada uma delas desempenhará. A qualidade de uma RS e o tempo necessário para realizá-la dependerão da habilidade, experiência e conhecimentos técnicos dos envolvidos [17] [19] [20] [21]. São ainda requeridos dos envolvidos itens como noções de estatística, domínio do tema que está sendo pesquisado, domínio da língua inglesa e, principalmente, conhecimento do processo de RS.

De acordo com Carver et. al [21], a realização de uma RS é difícil e demorada para pesquisadores considerados experientes, e é mais ainda para os novatos, por exemplo, estudantes de pós-graduação. Além disso, Riaz et al. [19] observam que a formulação das questões de pesquisa é igualmente desafiadora tanto para novatos quanto para pesquisadores experientes. Outros trabalhos [9] [22] destacam também a dificuldade em se definir o alcance e relevância das questões de pesquisa.

Observa-se que grandes desafios podem ser enfrentados por pesquisadores novatos, principalmente em função de sua limitada experiência com RSs e, muitas vezes, com o tópico de pesquisa em questão [17]. Para que pesquisadores novatos consigam planejar melhor suas RSs, eles devem buscar orientação com pesquisadores mais experientes, como os professores orientadores. Além disso, o uso de modelos de formulários e planilhas para, por exemplo, a seleção e extração de dados, bem como a leitura de publicações apresentando RSs são de grande valia. Por fim, observa-se que a grande maioria das RSs encontradas na área de ES foram conduzidas com uma forte atuação de alunos de pós-graduação, mas que contou com o suporte de pesquisadores experientes da área [20].

Mesmo utilizando-se de um processo bem definido, RSs são trabalhosas e requerem grande esforço dos envolvidos [9] [19] [21] [23], o que faz com que o uso de ferramentas de software de apoio à sua execução se torne interessante. Marshall, Brereton e Kitchenham [24] [25] [26] identificaram as ferramentas que têm sido utilizadas para automatizar parte ou a totalidade do processo de RS no domínio de ES. Dentre as ferramentas que auxiliam parte do processo de RS, destacam-se a DBpedia [27], um repositório de estudos para auxiliar a seleção automática de estudos; a Revis (*Systematic Literature Review Supported by Visual Analytics*), que apoia a atividade de seleção de estudos por meio da aplicação da mineração visual de texto (*Visual*

Text Mining - VTM) [28]; a UNITEX [29], que dá suporte à identificação automática de resultados; e a PEx-Graph [30], que disponibiliza representações gráficas dos resultados de RSs.

Das ferramentas que apoiam todo o processo de RS, ou seja, desde a fase de planejamento, passando pelas atividades de seleção de estudos primários e extração de dados desses estudos na fase de condução, até a fase de sumarização dos dados, destacam-se a SLuRp (*Systematic Literature unified Review Program*) [31]; a StArt (*State of the Art through systematic review*) [32]; a SLR-Tool (*A Tool for performing Systematic Literature Reviews*) [33]; e a SLRTOOL (*A Tool to Support Collaborative Systematic Literature Reviews*) [34]. A Tabela 1.1 apresenta as principais funcionalidades dessas ferramentas.

Tabela 1.1 Ferramentas de apoio à RS [24] [25] [26].

Ferramentas	SLuRp	StArt	SLR-Tool	SLRTOOL
Funcionalidades				
Desenvolvimento do protocolo	✓	✓	✓	
Supor te às buscas automatizadas	P			
Seleção de estudos	✓	✓	✓	P
Avaliação de qualidade	✓		P	P
Extração de dados	✓	✓	✓	✓
Análise computacional de texto	✓	✓		
Metanálise	P			
Escrita de relatório	P	P	P	
P → Atende parcialmente.				

Além das ferramentas que apoiam todo o processo de RS, ferramentas mais gerais, tais como os gerenciadores de referências bibliográficas, também podem ser utilizadas para auxiliar na condução de RSs. Alguns exemplos são o EndNote¹ (comercial e disponível para as plataformas Windows e Mac OS X); JabRef² (gratuito e disponível para as plataformas Windows, Mac OS X e Linux); e Mendeley³ (gratuito na versão básica e disponível para as plataformas Windows, Mac OS X e Linux). Essas ferramentas facilitam o gerenciamento dos estudos primários durante a condução da RS, por exemplo, possibilitando a classificação em ordem alfabética de título, nome do autor ou número de identificação, inclusive dando importante suporte para a remoção de estudos primários duplicados.

¹ <http://endnote.com>

² <http://jabref.sourceforge.net>

³ <https://www.mendeley.com>

Por último, mas não menos importante, diversos recursos materiais devem ser considerados para auxiliar a execução das RSs. De maneira geral, esses recursos são: computadores; acesso à Internet; acesso às bases bibliográficas (por exemplo, *IEEE Xplorer*, *ACM Digital library* e *ScienceDirect*), inclusive com acesso ao texto completo dos estudos primários; materiais de escritório em geral; e espaço para trabalho e/ou reuniões da equipe que conduzirá a RS.

Questão 5: Qual a confiabilidade de uma RS?

Em função do aumento do número de RSs conduzidas na ES, tornou-se relevante compreender a confiabilidade das revisões publicadas. Entende-se por confiabilidade o fato de duas revisões conduzidas independentemente e sobre um mesmo tópico de pesquisa chegarem às mesmas conclusões [35]. Nesse contexto, uma investigação sobre a confiabilidade de RSs como um método de pesquisa mostrou que duas RSs independentes são capazes de gerar as mesmas conclusões, desde que sejam conduzidas por pesquisadores com experiência no processo de RS e no tópico sendo investigado [36].

Quanto à reproduzibilidade das RSs, em particular, na atividade de seleção de estudos primários, observa-se que pesquisadores considerados novatos não vão necessariamente selecionar os mesmos estudos [37]. Portanto, não é possível garantir reproduzibilidade com relação à realização dessa atividade por novatos. Na mesma linha, um estudo de caso com alunos de segundo ano da graduação em Ciência da Computação mostrou que RSs podem ser realizadas por novatos, apesar de claramente desafiadora e demorada e, nesse caso, aconselha-se que sejam executadas em grupos [38]. Para esses novatos, a fase de condução, incluindo a atividade de seleção de estudos, é mais problemática do que a fase de planejamento da revisão. Por fim, para que a reproduzibilidade seja alcançada, pesquisadores experientes devem estar envolvidos na execução das RSs.

Questão 6: Quais são as vantagens, desvantagens e limitações de RSs?

Embora a revisão informal da literatura tenha importância na condução de uma pesquisa, principalmente quando realizada com subsídios de boas práticas, ela apresenta como principal desvantagem a falta de rigor científico, o que pode ser influenciada por vieses durante a investigação. Por outro lado, a condução de uma revisão com base em uma abordagem sistemática, como é o caso de RSs e MSs, resulta em um processo controlado e rigoroso, garantindo que a mesma seja passível de auditoria, repetível e imparcial aos interesses dos envolvidos. Outros benefícios de uma RS são:

- Uma RS bem planejada evita vieses na sumarização dos estudos primários;

- Uma RS permite responder questões de pesquisa que não são possíveis de serem respondidas por um único estudo primário;
- Uma RS permite identificar com mais precisão lacunas na pesquisa atual e, como consequência, identificar possíveis linhas de pesquisa para investigação futura;
- Os resultados dos estudos selecionados podem ser utilizados para comparação com os resultados obtidos em um trabalho em desenvolvimento;
- Uma RS pode ajudar a avaliar hipóteses teóricas que possivelmente nunca foram ou nunca poderiam ser testadas por meio de estudos primários;
- Uma RS pode beneficiar a indústria de software no apoio à tomada de decisões relacionadas a o que será adotado no futuro. Por exemplo, profissionais podem usar resultados de uma RS para compreender a eficácia e eficiência de um método ou tecnologia específica, ou então, apontar o que funciona e o que não funciona quando um método é aplicado sob diferentes circunstâncias.

RSs também apresentam dificuldades e limitações:

- A execução de uma RS requer mais esforço do que uma revisão informal, por exemplo na fase de planejamento, que inclui a definição do protocolo, não necessariamente definido na revisão informal;
- Atividades como a formulação das questões de pesquisa, a escrita da *string* de busca e a definição das bases de busca são consideradas difíceis;
- A seleção dos estudos é uma das principais dificuldades. Pesquisadores normalmente determinam a relevância dos estudos primários com base em seu título e resumo. No entanto, tanto o título quanto os resumos de muitos estudos não apresentam informação suficiente para essa finalidade. Outras partes dos estudos devem então ser lidas para determinar a relevância dos estudos para a pesquisa em questão;
- Nem sempre os estudos estão disponíveis para livre acesso (no caso, o texto completo), o que pode causar atrasos na execução da RS; e
- A avaliação da qualidade dos estudos primários, bem como a extração de dados para responder às questões de pesquisa são um desafio, uma vez que muitos estudos não apresentam em mais detalhes o trabalho realizado.

Considerações finais

Baseado na fundamentação teórica e práticas provindas da área de Medicina, a RS tem sido largamente aplicada nos últimos anos como um importante método de pesquisa na área de ES. Uma gama de RSs foram publicadas em importantes conferências e periódicos de ES e de suas

subáreas. Bons exemplos de RSs podem ser encontradas em [39] [40] [41] [42] [43]. Vale ainda destacar que RSs podem ser utilizadas como um método complementar às revisões informais, uma vez que cada qual apresenta suas vantagens, dificuldades e limitações. Em resumo, os principais pontos de aprendizagem deste capítulo foram:

O processo de RS teve origem na área da Medicina e tem despertado atenção de pesquisadores de diferentes áreas do conhecimento, a exemplo da Ciência da Computação e, mais especificamente, da ES;

- A RS é um dos principais métodos de pesquisa para sumarizar evidências na ESBE;
- Uma RS deve ser conduzida por meio de um processo de pesquisa sistemático, rigoroso e metodologicamente bem definido;
- O sucesso de uma RS depende diretamente da disponibilidade de um conjunto de recursos, a saber recursos humanos, de software e material; e
- RSs apresentam diversos benefícios, mas também dificuldades e limitações para a sua execução.

2. Protocolo da revisão sistemática

Sandra Camargo Pinto Ferraz Fabbri

Fábio Roberto Octaviano

Elis Cristina Montoro Hernandes

Toda RS deve ser realizada de acordo com um plano pré-definido. Esse plano é chamado de protocolo da revisão e formalizará todo o processo para a execução da RS. O protocolo contém, entre outros, itens como questões de pesquisa, estratégia de busca, critérios por meio dos quais os estudos serão avaliados para inclusão ou exclusão da revisão, e estratégias para seleção, extração e sumarização dos dados. Neste capítulo são apresentados cada um dos itens que compõem o protocolo. É destacada também a necessidade de realizar a avaliação do protocolo.

Questão 1: O que é o protocolo?

O protocolo é um documento criado durante a fase de planejamento da RS. O principal objetivo do protocolo é reduzir os vieses que podem ocorrer durante a execução de uma RS, definindo-se estratégias, critérios e formulários que serão seguidos pelos pesquisadores que realizarão a revisão. O preenchimento correto do protocolo direciona os pesquisadores nas próximas fases do processo de RS e permite que as estratégias e critérios adotados sejam compartilhados e interpretados por outros pesquisadores que se interessam pelo tópico investigado. O conteúdo do protocolo facilita a reproduzibilidade de uma RS, o que é de fundamental importância no próprio conceito de uma revisão. Presume-se que um pesquisador possa reproduzir todo o processo com base nas informações contidas no protocolo.

O protocolo é dividido em cinco seções: (1) Informações Gerais -- contém itens como título da RS, pesquisadores que conduzirão a RS, descrição da revisão e seus objetivos; (2) Questões de pesquisa -- contém itens como questões de pesquisa primárias e secundárias; (3) Identificação de estudos -- contém itens como palavras-chave, *strings* de busca, critérios de seleção das fontes de busca, lista das fontes de busca e a estratégia de busca; (4) Seleção e avaliação de estudos -- contém itens como critérios de inclusão e exclusão, estratégia para seleção dos estudos e avaliação da qualidade dos estudos; e (5) Síntese dos dados e apresentação dos resultados -- contém itens como estratégia de extração e sumarização dos dados e estratégia de publicação dos resultados. Opcionalmente, pode-se descrever o cronograma para a execução da revisão. A Tabela 2.1 mostra uma breve descrição dessas seções e seus respectivos itens.

Devido à importância do protocolo e com base em lições aprendidas ao executar nossas RSs, assim como Brereton et al. [9], sugerimos recomendações aos pesquisadores para a criação do protocolo:

- Estar atento à possibilidade de revisar as questões de pesquisa à medida que o entendimento do problema aumente no decorrer da revisão;
- Fazer com que todos os pesquisadores envolvidos na revisão participem da criação do protocolo;
- Realizar um teste piloto (simulação das fases) considerando alguns estudos primários para averiguar se não há necessidade de modificações, uma vez que a elaboração do protocolo é um processo iterativo e que pode requerer diversas revisões; e
- Realizar um mapeamento inicial da área de pesquisa antes de iniciar a revisão propriamente dita, o que pode auxiliar na criação do protocolo.

Cada uma das seções do protocolo, com seus respectivos itens (descritos na Tabela 2.1), são explicados com mais detalhes nas próximas questões.

Tabela 2.1: Seções e itens que compõem o protocolo.

Informações gerais	
Título	<título da RS -- deve ser um título sugestivo que caracterize o tema investigado>
Pesquisadores	<nome dos pesquisadores envolvidos>
Descrição	<explicação e justificativas que levam à condução da RS>
Objetivos	<declaração objetiva e precisa daquilo que se pretende investigar>
Questões de Pesquisa	
Questão de Pesquisa	<objetivo descrito no formato de questões de pesquisa>
Identificação de estudos	
Palavras-chave	<palavras que caracterizem o tema investigado e que sejam efetivas na busca de estudos relacionados ao tema>
Strings de busca	<conjunto de palavras e termos referente ao tema de pesquisa conectados por operadores lógicos AND e OR>
Critérios de seleção das fontes de busca	<critérios que determinem como selecionar as fontes de busca nas quais os estudos serão buscados>
Lista das fontes de busca	<bases bibliográficas e outras fontes que satisfazem a critérios e, portanto, serão utilizadas para a realização das buscas>
Estratégia de busca	<procedimento que será adotado para realizar a busca nas fontes selecionadas>
Seleção e avaliação de estudos	
Critérios de inclusão e exclusão dos estudos	<critérios que serão utilizados para incluir os estudos relevantes e para descartar aqueles que não são relevantes para aquela RS>
Estratégia para seleção dos estudos	<estratégia que será adotada para selecionar os estudos nas etapas de seleção do processo de RS>
Avaliação da qualidade dos estudos	<estratégia e critérios que caracterizem a qualidade individual dos estudos>

Síntese dos dados e apresentação dos resultados	
Estratégia de extração de dados	<estratégia de como as informações de cada estudo serão extraídas e os respectivos dados de extração>
Estratégia de sumarização dos dados	<estratégia de como os dados extraídos dos estudos serão sumarizados e o tipo de análise que será realizada>
Estratégia de publicação	<estratégia de como os resultados serão disponibilizados aos interessados>

Questão 2: Qual é o objetivo e itens relacionados à Seção *Informações gerais*?

O objetivo da seção *Informações gerais* é documentar os dados globais sobre a RS. Assim, contém itens como título da RS, pesquisadores envolvidos, descrição e objetivos da RS. O título da RS deve representar o tópico de pesquisa que está sendo investigado. Exemplo: *Uma revisão sistemática sobre o desenvolvimento de sistemas robóticos orientados a serviços* ou *Uma revisão sistemática sobre ferramentas de apoio à condução de revisão sistemática*. Além do título, outras informações que devem ficar registradas são os pesquisadores envolvidos na RS, por meio de seus nomes e afiliação e qualquer outra informação que se julgue relevante.

Também deve ser documentada uma descrição da RS, na qual conste uma explicação e justificativas que levam à sua realização. Essa justificativa é importante, pois pode ajudar a entender melhor o contexto associado à RS, facilitando inclusive o entendimento da análise dos dados. Para sintetizar essa informação deve-se estabelecer também o objetivo da RS, que é outro item das informações gerais da RS.

Questão 3: Qual é o objetivo e itens relacionados à Seção *Questões de pesquisa*?

O objetivo da seção *Questões de pesquisa* é descrever as questões referentes ao tópico de pesquisa abordado e que devem ser respondidas. As questões de pesquisa podem ser primárias ou secundárias. A questão primária corresponde, na sua essência, ao objetivo descrito na Seção *Informações Gerais*, mas redigido no formato interrogativo, caracterizando o tópico de pesquisa que está sendo investigado. A questão primária pode ter uma ou mais questões secundárias que, no caso de existirem, devem ressaltar particularidades do tópico investigado.

As questões de pesquisa guiam toda a condução da revisão e, por isso, a elaboração das questões de pesquisa é considerada a parte mais importante de qualquer RS [7]. É com base nas questões de pesquisa que a atividade de busca e seleção de estudos primários deve ser realizada, ou seja, as questões auxiliam na identificação dos estudos primários que devem ser considerados relevantes (estudos aceitos) para aquela RS, uma vez que o objetivo é respondê-

las da forma mais completa possível. A atividade de extração de dados dos estudos considerados relevantes também deve ser totalmente orientada pelas questões de pesquisa, pois é importante que se extraiam todos os dados para que as questões possam ser respondidas adequadamente. As questões de pesquisa também influenciam diretamente a atividade de análise e sumarização dos dados extraídos dos estudos, uma vez que eles precisam ser sintetizados de forma que as questões possam ser respondidas pelos pesquisadores. Assim, para o sucesso de uma RS, é vital que as questões de pesquisas sejam elaboradas corretamente.

Uma questão de pesquisa pode ser estruturada seguindo o conjunto de critérios PICO, derivados da Medicina [44]: P (população, paciente ou problema), I (intervenção), C (comparação, controle ou comparador) e O (resultados, do termo em inglês *outcomes*). Na Medicina, uma questão de pesquisa clínica é considerada adequada se pode ser estruturada em:

- População, ou seja, as pessoas ou pacientes afetados pelas intervenções, descrevendo as características mais importantes dos pacientes, o principal problema, doença ou suas condições. Em muitos casos, o sexo, idade ou raça de um paciente pode ser relevante para o diagnóstico ou tratamento de uma doença;
- Intervenção, que normalmente é uma comparação entre tratamentos alternativos. Pode indicar qual fator prognóstico ou a exposição que é considerada, o que se pretende fazer para o paciente (prescrever uma droga, pedir um teste, marcar uma cirurgia, etc);
- Comparação ou controle, que procura mencionar as principais alternativas para comparar com a intervenção, como por exemplo, se está tentando se decidir entre duas drogas, uma droga (intervenção) e nenhuma medicação ou placebo (comparação). Em alguns casos, esse critério pode ser omitido, pois nem sempre precisa haver uma comparação específica em questões clínicas; e
- Resultados, que procuram listar o que se pretende realizar, medir, melhorar ou afetar em relação à população (aliviar ou eliminar os sintomas, reduzir o número de eventos adversos, etc).

Pesquisadores da Universidade Médica Weill Cornell⁴ mostram um exemplo de uso dos critérios PICO para criar uma questão de pesquisa da área de Medicina. Com base nesses critérios, exibidos na Tabela 2.2, eles elaboraram a seguinte questão de pesquisa: *Em*

⁴ <http://qatar-weill.cornell.edu/elibrary/index.html>

pacientes idosos, os inibidores de ECA (Enzima Conversora da Angiotensina) são mais eficazes do que os betabloqueadores no controle da pressão arterial e na minimização dos efeitos adversos?

Tabela 2.2: Critérios PICO referentes a um exemplo da área médica.

População/Pacientes	hipertensão arterial, idosos
Intervenção	betabloqueadores
Comparação	inibidores de ECA
Resultado	reduzir a pressão arterial, minimizar os efeitos adversos

O grande desafio, quando se trata de outra área que não a Medicina, é adaptar o conjunto de critérios PICO. Para a área de ES, duas propostas de adaptações existentes na literatura são apresentadas abaixo da seguinte forma: (K) Kitchenham e Charts [7] e (B) Biolchini et al. [6]:

- População:

- (K): um papel específico da ES (por exemplo, testador, gerente); uma categoria específica de engenheiro de software (por exemplo, novato, experiente); uma área de aplicação (por exemplo, sistemas de TI, sistemas de controle); um grupo da indústria (por exemplo, companhia de telecomunicação, pequena empresa).

- (B): grupo da população que será observado pela intervenção (por exemplo, publicações que tratam do assunto investigado).

- Intervenção:

- (K): método, procedimento, tecnologia ou ferramenta que serão investigados.

- (B): aquilo que será investigado.

- Comparação ou controle:

- (K): método, procedimento, tecnologia ou ferramenta com a qual a intervenção deve ser comparada.

- (B): parâmetro de referência ou um conjunto de dados iniciais que o pesquisador já possua.

- Resultados:

- (K): os resultados estão relacionados aos fatores considerados importantes para caracterizar o que está sendo investigado e que são de interesse àqueles que farão uso da RS.

- (B): correspondem ao campo *Efeito*, no qual se relatam os tipos de resultados esperados ao final da RS.

Em seguida, apresentam-se dois exemplos de uso do PICO para a estruturação de questões de pesquisa no contexto da ES. O primeiro exemplo é de um trabalho de Biolchini et al. [6], cuja questão de pesquisa é: *Quais iniciativas têm sido propostas para avaliar a descrição de processos no contexto de ES?* Na Tabela 2.3 essa questão é estruturada segundo o conjunto de critérios PICO.

Tabela 2.3: Critérios PICO na área de ES - descrição de processos.

População	publicações relacionadas à definição e revisão de processos de software
Intervenção	avaliação de descrição de processos de software
Comparação	não se aplica
Resultado	identificação de iniciativas relacionadas à revisão e verificação

Outro exemplo de uso do PICO pode ser encontrado no trabalho de Oliveira et al. [45]. Os autores criaram as seguintes questões de pesquisa:

- *QP1: Como a orientação a serviços tem sido aplicada no desenvolvimento de sistemas robóticos?*
- *QP2: Qual a forma mais comum de interação entre sistemas robóticos orientados a serviço?*
- *QP3: Qual tecnologia de implementação tem sido a mais usada para desenvolver sistemas robóticos orientados a serviço?*
- *QP4: Quais são os ambientes de desenvolvimento e as ferramentas que dão suporte ao desenvolvimento de sistemas robóticos orientados a serviço?*
- *QP5: Arquitetura Orientada a Serviços (SOA) é aplicável a qualquer tipo de robô? e*
- *QP6: Como a engenharia de software tem sido aplicada no desenvolvimento de sistemas robóticos orientados a serviço?*

Na Tabela 2.4 essas questões são estruturadas segundo o conjunto de critérios PICO.

Como mencionado anteriormente, adaptar o conjunto de critérios PICO para a ES não é uma tarefa trivial. Observe que no nosso primeiro exemplo [6], a População foi definida como sendo as publicações relacionadas à temática da RS -- publicações relacionadas à definição e revisão de processos de software. Ou seja, os autores consideraram a População como sendo o grupo que será observado pela intervenção. Já no segundo exemplo [45], a População foi definida como sendo o público alvo interessado nos resultados da RS. No entanto, se

considerarmos a definição de Biolchini [6], a população do trabalho de Oliveira et al. poderia ser definida como os estudos disponíveis na literatura sobre SOA, que é o tema investigado por essa RS. Diferentemente da ES, na Medicina, derivar os critérios PICO é muito mais intuitivo, pois a População está diretamente relacionada com pessoas, que são os pacientes observados em um estudo.

Tabela 2.4: Critérios PICO na área de ES - orientação a serviço e sistemas robóticos.

População	Pesquisadores e desenvolvedores de sistemas robóticos que usam ou estão interessados no uso de SOA como base para o projeto de seus sistemas
Intervenção	sistemas robóticos orientados a serviço, seus projetos e os métodos usados em seu desenvolvimento
Comparação	não se aplica
Resultado	visão geral dos estudos propostos na literatura que reportam o uso de SOA, os domínios para os quais os robôs foram construídos, e as estratégias, métodos e tecnologias que deram suporte ao seu desenvolvimento

Em resumo, as questões de pesquisa devem ser cuidadosamente formuladas, pois elas vão direcionar todas as outras atividades da RS, sendo consenso na literatura. Dybå, Dingsøyr e Hanssen [46] salientam que é a partir das questões de pesquisa que são derivadas as palavras-chave que serão utilizadas na busca dos estudos. Brereton et al. [9] salientam que as questões de pesquisa são o ponto mais crítico da RS, pois é a partir delas que são construídas as *strings* de busca e que é direcionada à atividade de extração de dados. Além disso, esses autores consideram que as questões de pesquisa não devem ser alteradas após a avaliação do protocolo. Em decorrência disso, eles sugerem que seja feita um teste piloto para refinamento do protocolo, em especial, para que as questões de pesquisa caracterizem adequadamente o objetivo da RS. Segundo Staples e Niazi [47], como as questões de pesquisa direcionam toda RS, é importante que as questões de pesquisa estejam precisamente definidas, pois isso trará impacto direto no esforço e duração da RS.

Para avaliar se uma questão de pesquisa está correta, Kitchenham e Charters [7] sugerem as seguintes verificações:

- Se a questão elaborada é significativa tanto para quem atua diretamente com o tópico da pesquisa quanto para os pesquisadores da área, ou seja, se ela reflete tanto os anseios de quem atua quanto quem pesquisa sobre o tópico abordado;
- Se a questão de pesquisa elaborada acarreta em mudanças nas práticas existentes na área do tópico de pesquisa ou se contribui para aumentar a confiança nas práticas atuais da área no que se refere ao tópico de pesquisa;
- Se a questão de pesquisa elaborada auxilia na identificação de discrepâncias entre o que se acredita/espera sobre o tópico de pesquisa e o que realmente ocorre na prática, ou seja, confrontar a expectativa com a realidade;
- Como pode haver questões de pesquisa que são apenas de interesse acadêmico, ou seja, de pesquisadores pertencentes às instituições de ensino, espera-se que essas

questões de pesquisa auxiliem na identificação de lacunas que levem a atividades de pesquisa complementares; e

- Para auxiliar na elaboração das questões, Staples e Niazi [47] recomendam limitar o escopo da RS por meio de questões de pesquisa estruturadas de maneira clara e objetiva. Uma das formas de se fazer isso é aplicar o conjunto de critérios PICO.

Questão 4: Qual é o objetivo e itens relacionados à Seção Identificação de estudos?

O objetivo da seção *Identificação de estudos* é documentar a estratégia de como conduzir a busca pelos estudos relevantes. Assim, contém itens como palavras-chave, *strings* de busca, critérios de seleção das fontes de busca, lista das fontes de busca e a estratégia de busca.

As palavras-chave são termos presentes nas questões de pesquisa e que, consequentemente, representam o objetivo da RS. Uma vez determinados esses termos, é importante que se identifiquem seus sinônimos (termos alternativos), abreviações, grafias alternativas e plural. A *string* de busca é a combinação das palavras-chave e termos relacionados usando operadores lógicos de tal forma que a maior quantidade de estudos seja encontrada nas bases bibliográficas.

As fontes de busca são os locais em que serão feitas as buscas pelos estudos relacionados ao tema investigado. Estabelecer fontes apropriadas é um ponto importante para que estudos relevantes não fiquem inacessíveis simplesmente por causa das fontes definidas. Algumas sugestões de critérios que podem ser estabelecidos para determinar quais fontes devem ser pesquisadas durante a RS são, por exemplo: bases bibliográficas às quais se tem acesso; bases mencionadas por mapeamentos e RSs publicadas por outros pesquisadores; e bases que indexam fóruns de discussão do tema investigado.

Vale ressaltar que a inserção manual de estudos não deve ser descartada em um processo de RS e, assim, uma fonte que não pode ser esquecida são as bibliotecas físicas, com suas publicações em papel, pois pode haver estudos relevantes que não estejam digitalizados e/ou indexados pelas bases bibliográficas.

A estratégia de busca é a forma com que os estudos serão procurados nas fontes estabelecidas. É importante descrever no protocolo como a busca será conduzida nas fontes para que, no caso de se querer repetir ou atualizar a revisão, consiga-se proceder da mesma forma. Assim, nessa seção do protocolo, o pesquisador deve explicar o procedimento que ele pretende adotar para fazer a busca, de forma que não haja dúvidas. Detalhes sobre a estratégia de busca de estudos são tratados no Capítulo 3.

Questão 5: Qual é o objetivo e itens relacionados à Seção Seleção e avaliação de estudos?

O objetivo da seção *Seleção e avaliação de estudos* é documentar como os estudos relevantes para responderem às questões de pesquisa da RS serão selecionados e avaliados. Assim, contém itens como critérios de inclusão e exclusão, estratégia para seleção dos estudos e a avaliação da qualidade dos estudos.

Critérios de inclusão e exclusão são critérios definidos pelos pesquisadores para guiar a seleção de estudos primários que sejam relevantes para aquela RS. Os critérios de inclusão servem para indicar por qual ou quais critérios um estudo foi incluído na revisão, ou seja, considerado relevante para a revisão. Da mesma forma, os critérios de exclusão servem para indicar por qual ou quais critérios um estudo foi excluído da revisão, ou seja, considerado não relevante para aquela mesma RS.

Uma vez definidos os critérios de inclusão e exclusão, os pesquisadores devem definir a estratégia para seleção dos estudos. A estratégia consiste tanto em estabelecer as etapas da seleção (seleção inicial, seleção final e revisão da seleção), como também a maneira como os critérios de inclusão e exclusão deverão ser aplicados durante essas etapas, ou seja, é preciso definir se para aceitar um estudo basta que ele satisfaça um critério de inclusão ou alguns dos critérios, ou, eventualmente, todos os critérios.

O objetivo principal ao utilizar critérios de qualidade é avaliar aspectos metodológicos dos estudos. Ao procurar avaliar a qualidade dos estudos primários por meio da aplicação de critérios de qualidade, o pesquisador busca aumentar a confiabilidade nos resultados que serão obtidos e também na generalização dos mesmos [7]. Uma maneira de se medir a qualidade de estudos primários é por meio da aplicação de *checklist*, ou seja, lista que contém itens que serão utilizados para avaliar a qualidade de cada estudo independentemente. Não há regras a serem seguidas na formulação do *checklist* de qualidade, mas ele pode se referir a como os dados daquele estudo foram analisados, como o estudo foi avaliado, como foram feitas as comparações de intervenções, como a população foi selecionada, ou seja, são questões que avaliam os aspectos metodológicos do estudo sob análise. Se os itens do *checklist* puderem ter escalas numéricas atribuídas a eles, então um resultado numérico de qualidade pode ser obtido, o que muitas vezes pode auxiliar na detecção de estudos com alta, média e baixa qualidade. Além da definição dos critérios de qualidade que serão aplicados e suas escalas, deve-se estabelecer o procedimento que será adotado para aplicar esses

critérios. A descrição do procedimento possibilita que a avaliação da qualidade seja reproduzida por outros pesquisadores.

Vale destacar que uma RS pode conter estudos primários que reportam resultados quantitativos e/ou qualitativos. É recomendado que se construa um *checklist* para avaliação da qualidade de estudos quantitativos e outro para a avaliação da qualidade de estudos qualitativos, visto que há uma grande diferença entre eles na maneira como os resultados são analisados e relatados.

Os resultados da avaliação da qualidade podem ser utilizados para auxiliar na seleção de estudos primários relevantes e/ou para ajudar na análise e sumarização dos dados. No primeiro caso, sugere-se que os dados de qualidade sejam coletados antes dos dados para preenchimento do formulário de extração, em formulário à parte, diferente do formulário de extração de dados, uma vez que estudos serão incluídos ou excluídos da revisão com base nos dados de qualidade obtidos. No segundo caso, os dados de qualidade podem ser coletados simultaneamente à atividade de extração de dados dos estudos, e os itens de qualidade e de extração podem estar no mesmo formulário. Detalhes sobre a estratégia para selecionar estudos são tratados no Capítulo 4.

Questão 6: Qual é o objetivo e itens relacionados à Seção *Síntese dos dados e apresentação dos resultados*?

O objetivo da seção *Síntese dos dados e apresentação dos resultados* é documentar como serão extraídos os dados dos estudos considerados relevantes para responderem às questões de pesquisa da RS e como esses dados serão sumarizados e divulgados. Assim, contém itens como estratégia de extração e sumarização de dados e estratégia de publicação.

O formulário de extração de dados contém os itens que deverão ser coletados dos estudos primários considerados relevantes para responder às questões de pesquisa. O formulário é formado por campos definidos pelos pesquisadores para que os dados a serem extraídos sejam registrados de maneira uniforme para todos os estudos e, em atividades posteriores, analisados e sumarizados.

Dados gerais sobre a RS, como o título, autores, ano, objetivos, entre outros, que fazem parte do formulário de extração de dados e são fáceis de serem coletados podem ser extraídos pelo pesquisador (para cada estudo primário) no momento da seleção final, ou seja, durante a leitura do texto completo. No entanto, o ideal seria uma nova leitura exclusivamente dedicada para refinar os dados extraídos e fazer o preenchimento dos demais campos do formulário.

Os campos de um formulário de extração podem permitir:

- Respostas descritivas: por exemplo, a universidade ou grupo de pesquisa envolvido no estudo primário;
- Respostas exclusivas de uma lista pré-definida: por exemplo, para a questão - *o estudo relata a técnica X?* - a opção exclusiva de resposta seria *sim* ou *não*; e
- Respostas não exclusivas de uma lista pré-definida: por exemplo, para a questão - *qual ou quais técnicas o estudo relata?* - as opções não-exclusivas de respostas seriam as técnicas possíveis de serem aplicadas (duas ou mais).

Kitchenham e Charters [7] afirmam que se os critérios de qualidade forem utilizados de maneira complementar aos critérios de inclusão e exclusão, então um formulário à parte deve ser criado referente à avaliação da qualidade. No caso dos critérios de qualidade serem utilizados para auxiliar na síntese dos dados, é possível a criação de apenas um formulário.

É importante destacar que como o formulário de extração é definido na etapa de planejamento de uma RS, pode ocorrer que, no momento da execução da atividade de extração de dados dos estudos relevantes surja a necessidade de inclusão de novos campos ou de alteração dos campos existentes. Isso se deve ao fato da identificação de alguma informação lida no texto completo de algum estudo que não era de ciência dos pesquisadores, ou não havia sido pensada por eles quando o protocolo foi criado. Assim, a alteração do formulário de extração no protocolo faz-se necessária, mas é preciso tomar cuidado com os estudos que já tiveram seus dados extraídos anteriormente à alteração. Em tal situação, recomenda-se que os estudos anteriores sejam revisitados para verificar se o campo criado ou alterado não precisa ser novamente preenchido para esses estudos [18].

A sumarização dos dados coletados pode ser qualitativa (ou descritiva) ou quantitativa, por meio da aplicação de métodos estatísticos (por exemplo, meta-análise) [7]. O importante é que durante essa atividade as questões de pesquisa sejam claramente respondidas. A sumarização qualitativa pode fazer uso de tabelas para exibir informações extraídas dos estudos, mostrando suas similaridades e diferenças, se os resultados são homogêneos ou heterogêneos, e as razões encontradas no caso de heterogeneidade, mas sempre com foco nas questões de pesquisa para não desviar do contexto da revisão.

A sumarização quantitativa também pode incluir tabelas para auxiliar na apresentação dos resultados. Quando possível, os resultados de diferentes estudos primários devem ser

mostrados de forma comparativa. Dados provenientes de campos do formulário de extração com listas de respostas pré-definidas, como Sim/Não, podem ser calculados com base em alguma medida estatística como probabilidade (risco) ou razão de chances (*odds ratio*), entre outras. A utilização de gráficos é de grande valor para um melhor entendimento dos resultados e, nesse contexto, um dos gráficos indicados por Kitchenham e Charters [7] é o gráfico de floresta (*forest plot*). No caso de uma RS que envolva tanto estudos primários que apresentam resultados qualitativos quanto estudos que apresentam resultados quantitativos, a síntese dos resultados deve ocorrer separadamente; além disso, deve-se investigar se, por exemplo, os resultados qualitativos podem ajudar a explicar os resultados quantitativos. Isso resultaria em uma integração e composição final dos resultados muito interessante [7].

Depois de sumarizados, os resultados de uma RS devem ser divulgados aos possíveis interessados. Assim, faz-se necessário que uma estratégia de divulgação dos resultados obtidos seja definida e documentada no protocolo. No meio acadêmico, a publicação em conferências e revistas científicas são as principais alternativas disponíveis, mas outras formas de divulgação podem ser consideradas, tais como relatórios técnicos, pôsteres, páginas web e livros. Em relação à formatação do relatório principal de uma RS, vale dizer que no caso de publicação em artigos ou revistas científicas, normalmente há um *template* a ser seguido, e o pesquisador deve publicar conforme a exigência do evento selecionado. Em outros meios, o formato é definido pelos próprios pesquisadores. Uma estrutura de conteúdo mínimo para divulgação dos resultados de RSs é apresentada em [7], e envolve os itens: Título, Autores, Resumo, Contexto, Questões de Pesquisa, Métodos de Revisão, Estudos Incluídos e Excluídos, Resultados, Discussão, Conclusão, Agradecimentos, Conflitos de Interesse, Referências e Apêndices. Detalhes sobre estratégias para extração/sumarização dos dados e publicação dos resultados são tratados no Capítulo 5.

Questão 7: Como avaliar o protocolo?

Como o protocolo é o documento que guia toda a RS, é de grande importância que, uma vez criado, ele seja devidamente avaliado antes da fase de condução da revisão. O ideal é que essa avaliação seja feita por pesquisadores mais experientes, sobretudo quando os pesquisadores que realizarão a revisão forem inexperientes em RS ou pesquisa em geral e não tiverem domínio sobre o tema a ser investigado. Por exemplo, no contexto acadêmico, alunos de mestrado ou doutorado deveriam solicitar que seus orientadores fizessem a avaliação de seus protocolos. No caso de pesquisas que possuem financiamento, uma possibilidade é que um

grupo de pesquisadores experientes e independentes seja contratado para avaliar o protocolo criado [7].

Um *checklist* para avaliar o protocolo, proposto pelo Centro de Revisões e Disseminações (CRD) da Universidade de York [48], contém perguntas como:

- Os objetivos da revisão estão claramente descritos?
- As questões de pesquisa definidas são claras e objetivas?
- As fontes a serem pesquisadas para identificar os estudos primários foram descritas? Houve alguma restrição?
- A *string* de busca é devidamente derivada das questões de pesquisa?
- Os critérios de inclusão e exclusão estão descritos claramente? Não haverá dúvidas em como aplicá-los?
- Está claro se os critérios de qualidade serão utilizados? Os critérios de qualidade e a forma como eles serão aplicados estão descritos claramente?
- Os campos definidos no formulário de extração de dados são suficientes para responder às questões de pesquisa?
- A maneira como os dados extraídos serão analisados e sumarizados está descrita? Ela é suficiente para responder às questões de pesquisa?

Uma vez que as respostas às perguntas forem positivas e satisfatórias, o protocolo pode ser considerado avaliado e a fase de condução da revisão poderá ser iniciada. Caso haja perguntas cujas respostas não sejam positivas, o protocolo deve ser revisado pelos pesquisadores e os itens relacionados a elas devem ser melhorados até que todas as perguntas sejam respondidas satisfatoriamente.

Considerações finais

A etapa de planejamento da RS consiste essencialmente na elaboração de um protocolo que é o documento que possibilita que a RS seja reproduzida posteriormente, pois nele ficam registrados todos os detalhes e estratégias adotadas pelos pesquisadores envolvidos. Assim, o foco deste capítulo foi apresentar os itens fundamentais do protocolo de uma RS. Em resumo, os principais pontos de aprendizagem deste capítulo foram:

- Elaborar um protocolo é a melhor maneira de ter sucesso na execução da RS;
- O protocolo direciona a execução das outras fases da RS;

- As questões de pesquisa é que sintetizam o foco da RS;
- Todos os itens do protocolo são importantes e especificar corretamente as estratégias a serem adotadas é que possibilita a reprodutibilidade da RS; e
- A avaliação do protocolo é essencial para melhorar a qualidade da RS.

3. Identificação de estudos

Fabiano Cutigi Ferrari
Lucas Bueno Ruas de Oliveira
Bento Rafael Siqueira
Cleiton Rodrigo Queiroz Silva

A condução de uma RS exige uma busca abrangente pelos estudos primários. A busca deve recuperar estudos que possam contribuir para a elaboração de respostas para as questões de pesquisa definidas no protocolo (mais detalhes sobre protocolo podem ser encontrados no Capítulo 2). A busca deve também evitar que uma grande quantidade de estudos irrelevantes seja recuperada das fontes consultadas. O processo de busca inicia-se com a avaliação das questões de pesquisa e a formulação da *string* de busca. A *string* deve então ser adaptada e aplicada em cada base bibliográfica selecionada. Adicionalmente, buscas manuais e complementares devem ser realizadas em fontes de estudos não consideradas pelas bases bibliográficas sendo utilizadas. A opinião de especialistas também pode contribuir para a identificação de fontes de estudos e estudos em particular. Os passos adotados na busca devem ser documentados para possibilitar futuras verificações, reproduções ou extensões da RS conduzida e são discutidos em detalhes ao longo deste capítulo.

Questão 1: Como realizar uma busca abrangente por estudos primários?

Como definido no Capítulo 2, uma estratégia de busca visa definir como os estudos relevantes que respondem às questões de pesquisa da RS serão identificados. A definição de uma estratégia abrangente de busca tem impacto substancial na qualidade dos resultados obtidos para uma RS. Uma estratégia de busca abrangente e bem executada leva a resultados que satisfazem os objetivos dos pesquisadores, refletindo apropriadamente o estado da arte do tópico de pesquisa em questão. Caso contrário, estudos primários relevantes para os objetivos pretendidos com a RS podem não ser identificados. Ressalta-se que a estratégia de busca começa a ser definida durante o planejamento da RS, que se concretiza com o estabelecimento do protocolo, cujos detalhes foram apresentados no Capítulo 2. Uma estratégia de busca inclui:

- Definição das fontes, ou seja, os locais nos quais os estudos serão procurados. Uma estratégia de busca abrangente deve contemplar estudos publicados em periódicos e anais de eventos científicos, como conferências, simpósios e *workshops*. Vale ressaltar que as buscas não devem ser limitadas às bases bibliográficas. É importante considerar também a literatura cinzenta, que inclui relatórios técnicos, estudos preliminares,

especificações técnicas, entre outros. Na RS realizada por Oliveira et al. [45], por exemplo, os pesquisadores avaliaram estudos publicados em periódicos, anais de eventos científicos e em relatórios técnicos;

- Definição dos métodos de busca, ou seja, como serão realizadas as buscas automáticas em bases bibliográficas, as buscas manuais, a aplicação do *snowballing*, a consulta aos especialistas, ou o uso combinado de um ou mais métodos; e
- Descrição da aplicação dos métodos de busca, por exemplo, para buscas automáticas, é preciso definir a *string* de busca, os campos nos quais a *string* será aplicada, como o título, resumo e palavras-chave, se há restrição quanto à data de publicação do estudo, etc.

Devido à iteratividade do processo de RS [18], a estratégia de busca pode ser melhorada - por exemplo, para tornar a busca mais (ou menos) abrangente - dependendo de resultados obtidos nas buscas piloto.

Pontos brevemente mencionados nesta questão e outros itens relevantes para a definição de uma estratégia eficaz e eficiente de busca serão discutidos em mais detalhes ao longo deste capítulo.

Questão 2: Quais são os passos para converter questões de pesquisa em uma estratégia abrangente de busca?

A formulação adequada das questões de pesquisa é crucial para a definição da estratégia de busca; sendo assim, as questões de pesquisa devem ser claras e objetivas [7] [9]. Toda questão de pesquisa, seja ela primária ou secundária, impacta a definição da estratégia de busca, que visa localizar o maior número possível de estudos relevantes [7] [49].

A eficácia e a eficiência da estratégia de busca podem ser medidas. Uma estratégia eficaz visa encontrar o máximo de estudos primários relevantes enquanto uma estratégia eficiente visa encontrar estudos primários relevantes com o menor custo possível. Zhang e Babar [50] definem o termo *gold standard* para o meio termo entre as duas estratégias e afirmam ser essa uma prática comumente usada em RSs. A precisão de uma estratégia de busca também pode ser medida [51] [52]. O pesquisador deverá optar por construir uma estratégia de busca ampla ou restritiva. Em uma estratégia ampla, um número maior de estudos será avaliado e muitos deles serão rejeitados. Esse tipo de estratégia localiza mais estudos primários relevantes e demanda maior esforço e tempo [51]. Já em uma estratégia restritiva, por sua vez, menos estudos são identificados. Esse tipo de estratégia tende a ser mais eficiente com

relação a custo e benefício [51], no entanto, ao realizar uma estratégia restritiva, o pesquisador assume o risco de não localizar algum estudo primário relevante [53].

Várias estratégias podem ser utilizadas para localizar os estudos primários de uma RS. Criar uma estratégia de busca robusta e abrangente não é uma tarefa trivial, principalmente para pesquisadores iniciantes. Apenas o uso da *string* de busca não é suficiente e o pesquisador deve explorar outras estratégias para obter um melhor resultado [7]. Consultar a lista de referências dos estudos primários relevantes e de revisões publicadas, periódicos, literatura cinzenta, relatórios técnicos, trabalhos em progresso, anais de conferências e a Internet podem ajudar o pesquisador na tarefa.

A *Pearl Growing* [54] é uma opção de estratégia para encontrar estudos primários. Para aplicar a *Pearl Growing*, o pesquisador deverá encontrar um conjunto inicial C1 composto por um ou mais estudos primários que serão aceitos em sua RS. Para localizar elementos para o conjunto C1, o pesquisador poderá utilizar um ou mais métodos de busca dentre os discutidos neste capítulo. Uma boa alternativa é o uso do *quasi-gold* como conjunto inicial, que será discutido mais adiante. Após realizar a leitura de todos os estudos primários do conjunto C1, o pesquisador deverá utilizar o conhecimento obtido nesses estudos primários para localizar novos estudos primários relevantes. Os novos estudos encontrados e os estudos primários já conhecidos formarão um novo conjunto C2. Com a leitura dos estudos primários contidos em C2 e repetindo o mesmo processo, um novo conjunto C3 será formado. O pesquisador deverá repetir o processo até que o próximo conjunto encontrado contenha exatamente os mesmos estudos primários do conjunto anterior, ou seja, até que nenhum novo estudo seja encontrado. Na ilustração da Figura 3.1, são representados os conjuntos de estudos construídos com uso da estratégia *Pearl Growing*. C1 representa o conjunto inicial e CN representa o último conjunto encontrado.

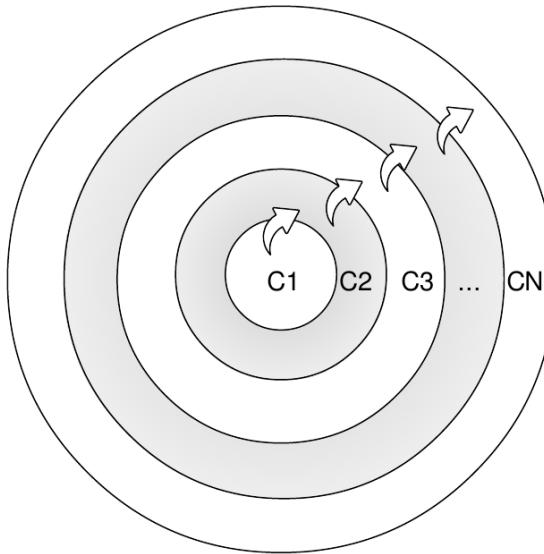


Figura 3.1: Representação em conjunto da estratégia Pearl Growing.

Pesquisadores inexperientes, ao aplicarem a estratégia *Pearl Growing*, podem iniciar suas pesquisas com buscas manuais na tentativa de explorar a literatura e as pesquisas existentes. Essas buscas permitem que os pesquisadores encontrem um conjunto inicial de estudos primários [50]. Com base nesse conjunto, o pesquisador deve identificar palavras-chave que poderão compor a *string* de busca. Após criada, é necessário garantir que a *string* de busca é adequada e que recupera bons resultados. Esse é um grande desafio. Segundo Fabbri et al. [18], a busca por estudos é uma atividade custosa e propensa a erros. A efetividade da busca está fortemente ligada à qualidade da *string* de busca e, por esse motivo, a *string* deve ser constantemente ajustada dentro do processo de execução da RS, uma vez que esse processo deve ser realizado iterativamente pelos revisores.

Questão 3: O que é e para que serve uma busca piloto?

A definição da *string* de busca é uma etapa fundamental para o sucesso de RSs. Contudo, a identificação de uma combinação de termos capaz de encontrar a maior quantidade possível de estudos primários relevantes de forma precisa requer experiência e conhecimento sobre a área de pesquisa [7]. Muitas vezes, os envolvidos na condução da RS não são especialistas ou estão realizando uma investigação de um tópico recente no qual novos termos podem surgir com relativa frequência [14]. Por isso, é importante realizar buscas preliminares visando avaliar a *string* mais adequada para cada RS [9]. Essa investigação inicial é conhecida como busca piloto.

Em uma busca piloto, pode-se utilizar revisões e MS conduzidos anteriormente e a opinião de especialistas, de forma a balizar a identificação dos principais termos e a construção de uma

string inicial [7]. Essa *string* pode, então, ser aprimorada com base em termos identificados em estudos primários relevantes retornados na busca em uma ou mais bases bibliográficas descritas no protocolo [18]. Estudos relevantes e previamente conhecidos sobre o tema da RS podem ser utilizados para formar um grupo de controle para avaliar a precisão da *string* de busca [7]. Não encontrar estudos do grupo de controle em nenhuma das bases bibliográficas é um forte indício de que a *string* ainda precisa ser aprimorada antes de ser aplicada de forma definitiva durante a condução da RS.

Uma outra estratégia importante para aprimorar a *string* durante a busca piloto é avaliar diferentes combinações de termos. Dessa forma, é possível verificar quais termos possuem real impacto na identificação de estudos primários relevantes e também eliminar termos que reduzam a precisão da busca e não agreguem estudos relevantes à RS. O trabalho de Oliveira et al. [45], por exemplo, identifica em sua busca piloto que o termo *SoC* retorna estudos primários relacionados à *System on Chip*, ao invés de *Service-oriented Computing*, quando utilizado em conjunto com termos como *Robot* e *Robotic System*. A busca piloto também permite identificar que estudos relacionados ao tema investigado apresentam o termo *Service-oriented*, já contido na *string*, sempre que o acrônimo *SoC* é utilizado, o que permite a remoção do mesmo sem prejuízo para a obtenção de estudos relevantes.

A identificação da *string* de busca mais adequada para cada RS é, de fato, uma atividade inherentemente iterativa [9]. Explorando essa característica, o trabalho de Fabbri et al. [18] propõe um processo que estende o aprimoramento da *string* de busca até a fase final de seleção de estudos primários. Com isso, termos identificados nos estudos primários lidos por completo durante a condução também podem ser utilizados para aprimorar a *string* de busca, o que pode garantir uma melhor cobertura da área de pesquisa investigada. Entretanto, vale lembrar que a alteração da *string* de busca durante a condução da RS implica em retornar à fase de planejamento, para reavaliação e atualização de outros elementos do protocolo, como o formulário de extração de dados.

Questão 4: A especificidade e a generalidade dos termos influenciam a recuperação de estudos?

Após a formulação das questões de pesquisas, é necessário definir uma estratégia para a obtenção de estudos, de forma a estabelecer um conjunto de evidências que permitam a elaboração das conclusões da RS [55]. Conforme discutido na Questão 3 (sobre a busca piloto), a criação de uma *string* de busca requer experiência e conhecimento sobre o assunto sendo

investigado. Nesse contexto, a definição de um conjunto adequado de termos para uma *string* de busca envolve um compromisso entre generalidade e especificidade.

Os termos utilizados na *string* podem ser classificados como controlados e não-controlados [56]. Termos controlados são conhecidos como descritores de assunto e podem ser obtidos a partir de glossários padrão, como os disponíveis em normas técnicas. Por exemplo, termos relacionados ao domínio de teste de software podem ser identificados na norma 610.12-1990-IEEE (*Standard Glossary of Software Engineering Terminology*) [57]. Por outro lado, termos não-controlados são palavras e seus sinônimos, variações de grafia, siglas e correlatos.

Por exemplo, o termo *modelagem* possui grafia distinta no inglês britânico (*modelling*) e no inglês americano (*modeling*), sendo que ambos são utilizados para a indexação de estudos primários nas bases bibliográficas. Isso acontece de maneira frequente, uma vez que os termos para a indexação são fornecidos pelos próprios autores dos estudos em partes como o título, resumo e palavras-chave. Na área de Ciência da Computação, em particular, não há um consenso na utilização desses termos fornecidos pelos autores [53]. Torna-se, assim, difícil a identificação de estudos com um conteúdo em particular. Inclusive pode haver vários estudos que tratam do mesmo tema, entretanto, indexados com termos diferentes.

Ao tentar expandir o conjunto de estudos primários da RS, deve-se sempre avaliar a sensibilidade e precisão dos termos utilizados na composição da *string* de busca [51]. A sensibilidade consiste em não recuperar um conjunto muito grande de estudos irrelevantes durante a aplicação da *string* nas bases bibliográficas. Por outro lado, a precisão refere-se a capacidade da *string* de busca em retornar o máximo de estudos relevantes durante uma consulta. Observa-se que a identificação da efetividade dos termos de uma *string* não é uma tarefa trivial, o que pode dificultar a avaliação da sensibilidade e da precisão de um conjunto de termos. Uma forma de mitigar tal complexidade é discutir os termos da *string* com especialistas [18]. Mais detalhes sobre o uso da opinião de especialistas são apresentados na Questão 6 deste capítulo.

É difícil mencionar que determinado termo traz estudos primários mais relevantes que outro, pois a relevância de um estudo primário é atribuída pelos objetivos e pelas questões de pesquisa. Entretanto, a saber, alguns termos podem ser mais específicos do que outros. No caso dos termos do trabalho de Oliveira et al. [45], por exemplo, o termo *Service Oriented* é mais específico do que o termo *software*. O limite da utilização desses termos deve estar relacionado não somente com as questões de pesquisa definidas para a RS, mas também com o contexto em que tal termo está inserido. Além disso, assim como introduzido anteriormente

na Questão 2, com a estratégia *Pearl Growing*, a cada fase do processo de pesquisa, o pesquisador evolui sua experiência com relação ao objeto de estudo. Desse modo, com o uso de uma abordagem iterativa, como a de Fabbri et al. [18], o conjunto de termos da *string* de busca evoluem.

Questão 5: Quais são as melhores fontes de estudos primários?

Durante a busca por estudos primários relevantes, deve-se adotar uma estratégia criteriosa, transparente e imparcial para a seleção das fontes de informação. Para isso, é necessário considerar fontes complementares, visando a mitigar possíveis limitações que induzam vieses ou limitem a identificação de estudos primários [7]. Um exemplo conhecido de viés na seleção de estudos é o fato de que resultados positivos tendem a ser mais facilmente publicados do que resultados negativos, o que pode vir a interferir na elaboração das respostas para as questões de pesquisa. Dessa forma, durante a definição do conjunto de estudos primários que será utilizado para a identificação de evidências, devem-se considerar tanto estudos primários indexados quanto estudos primários que contenham resultados não publicados [9].

Estudos primários indexados podem ser obtidos por meio da busca utilizando-se a *string* de busca nas diferentes bases bibliográficas. Essa atividade pode ser complementada por uma busca manual em outras fontes, tais como revistas e anais de eventos da área de pesquisa abordada pela RS. Além disso, é importante considerar como fonte adicional de evidência a lista de referência presente nos estudos primários incluídos [7]. Estudos primários não indexados podem, por outro lado, ser identificados em uma busca por relatórios técnicos, dissertações, teses, normas técnicas, registros de pesquisa ou revistas publicadas por empresas, tais como as organizadas pela *International Business Machines* (IBM)⁵. Nessa perspectiva, o contato com especialistas e pesquisadores do domínio pode facilitar na identificação de estudos primários que não foram publicados ou não estão indexados em bases bibliográficas. Uma discussão mais detalhada sobre busca manual e busca por estudos não indexados pode ser encontrada nas Questões 7 e 8 deste capítulo.

A escolha das bases bibliográficas a serem utilizadas na obtenção de estudos primários indexados deve considerar os conceitos de precisão e sensibilidade mencionados na Questão 4. Nenhuma base bibliográfica é capaz de retornar sozinha todos os estudos primários pertinentes a uma RS [9] [51]. Escolher um conjunto pouco abrangente de bases bibliográficas pode resultar na perda de estudos primários relevantes para responder às questões de pesquisa. Por outro lado, considerar todas as bases bibliográficas disponíveis para Ciência da

⁵ <http://www.research.ibm.com/journal>

Computação pode resultar em um conjunto muito grande de estudos fora do escopo da RS, o que leva a uma baixa precisão. Nesse contexto, os trabalhos de Brereton et al. [9], Kitchenham e Charters [7] e Dybå et al. [46] identificaram um conjunto de bases bibliográficas que podem ser úteis para a obtenção de estudos no domínio de ES. Esse conjunto de bases é listado na Tabela 3.1.

Tabela 3.1: Principais bases bibliográficas em Ciência da Computação.

Fonte de informação	Endereço online	Tipo
<i>IEEE Xplore</i>	ieeexplore.ieee.org	Base bibliográfica
<i>ACM Digital library</i>	dl.acm.org	Híbrida
<i>CiteSeer library</i>	citeseerx.ist.psu.edu	Motor de busca
<i>Google scholar</i>	scholar.google.com	Motor de busca
<i>Inspec</i>	www.theiet.org/resources/inspec	Motor de busca
<i>Engineering Village</i>	www.engineeringvillage.com	Motor de busca
<i>SpringerLink</i>	link.springer.com	Base bibliográfica
<i>Scopus</i>	www.scopus.com	Motor de busca
<i>ScienceDirect</i>	www.sciencedirect.com	Base bibliográfica
<i>Wiley online library</i>	onlinelibrary.wiley.com	Híbrida
<i>ISI Web of Science</i>	www.webofknowledge.com	Motor de busca

Bases bibliográficas, como a *SpringerLink*, retornam apenas estudos publicados pela própria editora. Por outro lado, motores de busca são capazes de retornar estudos indexados por diversas bases bibliográficas. Motores de busca, como a *Scopus* e o *Google Scholar*, por exemplo, indexam estudos publicados em algumas revistas e conferências da *ACM Digital library* e *IEEE Xplore*. Além disso, existem fontes de informação que indexam estudos primários publicados pela própria editora, mas também retornam estudos presentes em outras bases bibliográficas [58]. Esse tipo de fonte de informação pode ser denominada como híbrida, que é o caso da *ACM Digital library*.

É importante notar que, diferentemente de áreas como a Medicina, as bases bibliográficas disponíveis para a Ciência da Computação não são projetadas para apoiar a condução de RSs [9]. Estudos como os de Brereton et al. [9], Dybå et al. [46] e Dieste et al. [51] destacaram diversas limitações acerca das bases bibliográficas apresentadas na Tabela 3.1. Entre essas limitações, pode-se destacar:

- Recursos bibliográficos limitados: Algumas bases bibliográficas, tais como a *IEEE Xplore*, *SpringerLink* e *ScienceDirect*, não cobrem publicações de forma abrangente, limitando-se apenas a estudos publicados pela própria editora;

- Problemas com o algoritmo de busca: Parte das bases bibliográficas, como a *SpringerLink*, não permite limitar a busca considerando somente, por exemplo, título, resumo e palavras-chave [9] [46] [51]. Essa limitação reduz a precisão da busca e faz com que uma grande quantidade de estudos irrelevantes seja retornada, dificultando a realização da RS;
- Falha no reconhecimento de plurais: Em alguns casos, como no da *IEEE Xplore*, o plural de cada termo deve ser informado [51]. Por isso, antes da utilização de uma base bibliográfica, deve-se verificar a necessidade de especificar ou não termos no plural;
- Resumos ou textos de estudos primários incompletos: Algumas das bases bibliográficas, como a *IEEE Xplore*, *ACM Digital library* e *Scopus*, indexam estudos nos quais apenas o resumo está disponível. Essa limitação dificulta a avaliação e a utilização dos estudos durante a condução da RS [51]; e
- Uso de *strings* muito grandes: Em alguns casos, como na *IEEE Xplore*, bases bibliográficas impõem limitações quanto ao número de caracteres ou de termos. Essa limitação faz com que a *string* de busca precise ser dividida em partes menores, demandando múltiplas consultas e a retirada dos estudos repetidos obtidos por cada consulta.

Vale a pena destacar que essas limitações foram observadas no momento da escrita deste livro, podendo ser minimizadas futuramente. É importante considerar os pontos fortes e fracos de cada uma das bases bibliográficas sempre que se deseja estabelecer uma estratégia de busca abrangente, bem como ser capaz de identificar e mitigar possíveis limitações das bases consideradas na RS. Uma outra questão a se considerar é a sobreposição de conteúdo entre as bases bibliográficas, o que leva à obtenção estudos primários repetidos, conforme destacado por Oliveira et al. [45] e Dybå et al. [46]. Na RS de Oliveira et al. [45], por exemplo, foi identificado que a base bibliográfica *Scopus* já indexava todos os estudos retornados pela base bibliográfica *Engineering Village (IE Compendex)*, sendo a segunda pouco útil na obtenção de novos estudos. Para auxiliar na seleção das bases bibliográficas, Dieste et al. [51] propõem um conjunto de critérios:

- Abrangência: O conteúdo indexado pela base bibliográfica deve englobar um bom número de estudos relacionados à área de pesquisa;
- Atualização de conteúdo: A base bibliográfica deve indexar novos estudos regularmente;

- Disponibilidade de textos completos: Os textos completos dos estudos devem estar disponíveis na íntegra;
- Qualidade da busca: O mecanismo de busca da base deve permitir uma fácil adaptação da *string* de busca para os campos e comandos disponíveis para realizar a consulta na base bibliográfica;
- Qualidade dos resultados: As consultas na base bibliográfica devem retornar estudos de qualidade, que podem possuir evidências confiáveis;
- Versatilidade na exportação de resultados: A base bibliográfica deve possibilitar a exportação dos resultados obtidos na consulta, facilitando a execução de RSs, uma vez que RSs tendem a ter um grande número de estudos primários durante a etapa de seleção; e
- Usabilidade: A interface gráfica de acesso à base deve ser fácil de entender e operar, além de permitir a integração com aplicativos que facilitem o gerenciamento da bibliografia, tais como o *EndNote*⁶, *Mendeley*⁷ e *StArt*⁸. Outras ferramentas são citadas na Questão 4 do Capítulo 1.

Considerando os critérios apresentados, é possível obter um bom conjunto de estudos primários indexados e, ao mesmo tempo, realizar uma busca eficiente. Uma boa estratégia é combinar motores de busca abrangentes, como a *Scopus*, com outras bases mais específicas, como a *IEEE Xplore* e a *ACM Digital library*. A escolha das bases bibliográficas mais específicas deve levar em consideração eventos científicos e periódicos da área investigada que não são indexados pelos motores de busca. RSs conduzidas por Oliveira et al. [45] e Dieste et al. [51] reportam que a *Scopus* foi responsável pela indexação de aproximadamente 90% dos estudos. A utilização da *Scopus* também é indicada pelas diretrizes propostas por Kitchenham e Charters [7] para a condução de RSs. Contudo, vale lembrar que o impacto de cada base bibliográfica na obtenção de estudos primários pode variar de uma RS para outra. Por exemplo, a RS conduzida por Santos et al. [52] relata um menor impacto da *Scopus*, se comparado aos estudos supracitados.

Independentemente das bases bibliográficas utilizadas, vale lembrar que a escolha das fontes de informação deve ser definida de acordo com critérios imparciais e documentados no protocolo, de forma a permitir a auditoria e reproduzibilidade da RS.

Questão 6: Qual a importância e os riscos de consultar especialistas?

⁶ <http://endnote.com>

⁷ <https://www.mendeley.com>

⁸ http://lapes.dc.ufscar.br/tools/start_tool

A condução de uma RS demanda conhecimento prévio sobre a área de pesquisa investigada, uma vez que, ainda durante o protocolo, os envolvidos devem ser capazes de definir questões de pesquisa relevantes, critérios de inclusão e exclusão eficientes e não subjetivos, uma *string* de busca adequada e também o formulário de extração de dados. Entretanto, em muitos casos, os envolvidos em uma RS não possuem grande experiência sobre o tema, o que leva à necessidade de se considerar pesquisadores mais experientes e especialistas para aconselhamento durante o processo. Kitchenham e Charters [7], por exemplo, ressaltam a necessidade de alunos de mestrado e doutorado apresentarem e discutirem o protocolo da revisão com seus orientadores.

Especialistas da área sendo pesquisada podem fornecer apoio em todas as fases da RS. Durante a fase de planejamento, a opinião de especialistas pode ser fundamental para aumentar a qualidade do protocolo, fornecendo direções para a construção da *string* de busca e do grupo de controle de estudos primários [7] [9] [49]. Durante a condução, especialistas podem ser consultados para auxiliar na identificação de estudos primários não indexados (por exemplo, dissertações, teses, normas e relatórios técnicos) [7] e também na resolução de conflitos durante a seleção dos estudos [59]. O especialista pode, por exemplo, ajudar a ponderar se um dado estudo incluído por um revisor e excluído por outro será considerado na análise [9]. Por fim, durante a síntese e apresentação dos resultados, especialistas podem ser convidados a avaliar tanto o documento usado para a disseminação da RS, quanto a pertinência da análise realizada com relação às questões de pesquisa.

Mesmo considerando todos os benefícios da participação de especialistas durante uma RS, é importante estar ciente das possíveis ameaças à validade da revisão. Ainda que involuntariamente, especialistas podem influenciar na inclusão de determinados estudos primários em detrimento a outros. Em particular, ao sugerir estudos para o grupo de controle ou para a inclusão manual, o especialista pode ser influenciado por sua experiência prévia ou ponto de vista de pesquisa [60]. Dessa forma, é sempre importante justificar e documentar todas as decisões tomadas com base nas sugestões de especialista, de forma a permitir a auditoria e reproduzibilidade da RS.

Questão 7: Quais as diferenças entre busca automática e manual?

Uma busca automática é uma busca realizada com apoio computacional em bases bibliográficas. O pesquisador apenas fornece dados de entrada como parâmetros para a busca; porém, não exerce uma papel ativo na execução da mesma. A busca utilizando *string* de busca,

quando executada pelas bases bibliográficas de forma transparente ao usuário, é considerada automática [61].

Uma busca manual é a busca na qual o pesquisador executaativamente a busca. Consiste em executar uma sucessão de buscas menores em diferentes meios que indexam publicações bibliográficas [61] [62]. As buscas podem ser executadas em acervos, por exemplo, localizados na Internet, anais de conferências ou buscadores da Internet, visando localizar estudos primários relevantes. Para iniciar uma busca manual, o pesquisador deve primeiramente determinar quais tipos de estudo pretende localizar e delinear os melhores locais para encontrar os estudos relevantes. Feito isso, deverá coletar os estudos primários encontrados, documentando o processo. Segundo Zhang e Babar [50], uma busca manual é conduzida avaliando os estudos, um a um, nas fontes selecionadas pelo pesquisador. Existe uma série de desafios que tornam a busca manual suscetível a problemas. Ao realizar a busca manual, o pesquisador deve ter em mente que é muito complicado avaliar a abrangência da pesquisa, ou seja, é possível que muitos estudos relevantes sejam ignorados por não estarem acessíveis nos meios pesquisados [61]. Um dos principais objetivos da RS é assegurar que todo o processo seja replicável [7]. Ao realizar busca manual, o pesquisador deve levar em consideração que a estratégia empregada deve estar o mais próximo possível de um método sistemático e procedural para uma estratégia adequada [61]. Independentemente do pesquisador que execute a busca, ao realizar o procedimento de busca como descrito e documentado por outro pesquisador, um resultado equivalente deve ser encontrado.

Para buscas automáticas é relativamente mais simples documentar o processo, sendo necessário relatar os dados de entrada [6] [10], que na maioria dos casos é a própria *string* de busca. A atualização desse tipo de busca também é muito mais simples, sendo necessário apenas reexecutar as consultas nas bases bibliográficas e reavaliar os resultados. Para buscas manuais, basicamente, todas as buscas devem ser feitas manualmente [62].

Nota-se que, independentemente do tipo de busca realizado, vieses sempre poderão ocorrer. Na busca manual, a habilidade do pesquisador em executar o processo de busca pode influenciar nos resultados obtidos. A busca é um processo basicamente manual e sem diretrizes claras. Apesar de diversos problemas existentes em buscas manuais, existem indícios de que buscas manuais são benéficas e também mais efetivas para encontrar estudos primários relevantes [63]. No caso das buscas automáticas, conforme discutido na Questão 5, a escolha das bases equivocadamente pode comprometer os resultados. O conhecimento do

pesquisador dos termos a serem utilizados na busca também pode comprometer a qualidade desse tipo de busca.

Buscas manuais podem complementar os resultados da busca automática [61]. Seguindo as diretrizes para a execução de RSs em ES [7], o objetivo central é identificar toda pesquisa relevante existente para um tópico de interesse. Com base nessa prerrogativa, encontrar o máximo de estudos relevantes é vital para a execução adequada de RSs.

Segundo Zhang e Babar [50], diversos pesquisadores combinam buscas automáticas com buscas manuais. Muitas vezes, pesquisadores optam por iniciar suas pesquisas com buscas manuais na tentativa de mapear a literatura e as pesquisas existentes. Essas buscas permitem que os pesquisadores encontrem um conjunto inicial de estudos primários que posteriormente serão utilizados para construir uma estratégia de busca automática. Uma vez que o pesquisador detém conhecimento prévio sobre o tópico de pesquisa, uma estratégia de busca automática é construída e aplicada. Como resultado dessa estratégia, estudos são encontrados [54]. Uma excelente alternativa é aplicar novamente uma busca manual para avaliar se os resultados são abrangentes o suficiente para conter os resultados obtidos por buscas manuais. Não conter os resultados das buscas manuais sobre o mesmo tema pode significar problemas na estratégia automática e cabe ao pesquisador avaliar o que os novos resultados representam. Kitchenham e Brereton [63] recomendam utilizar em RSs uma combinação da busca manual com a busca automática. Utilizar apenas uma única estratégia de busca é aceitar o risco de inserir viés na pesquisa, em contrapartida, utilizar mais estratégias ocasiona maiores esforços e tempo.

Questão 8: Como localizar estudos na literatura cinzenta?

Assim como mencionado na Questão 5, o contato com especialistas e pesquisadores da área pode facilitar a identificação de estudos primários que não foram publicados ou não estão disponíveis de forma indexada. Uma maneira de se localizar esses estudos é a busca na literatura cinzenta.

A literatura cinzenta⁹ é definida como aquela que é produzida sobre todos os níveis governamentais, acadêmicos e industriais, que pode ser obtida nos formatos impresso (por exemplo, relatórios técnicos em bibliotecas físicas) e eletrônico (por exemplo, documentos da Internet). Além disso, essa literatura não é controlada por publicadores comerciais (por exemplo, *ACM Digital library*, *IEEE Xplore* ou *Elsevier*).

⁹ www.greylit.org

Em geral, incluem-se nessa categoria de literatura, não se limitando a esses, os seguintes tipos de materiais [64]: relatórios técnicos, estudos preliminares, especificações técnicas, documentos oficiais de órgãos específicos, dentre outros.

A literatura cinzenta serve como complemento de uma estratégia de busca. Assim, antes da execução da estratégia de busca, pode-se, por exemplo, utilizar a literatura cinzenta a fim de se obter uma visão geral sobre o tópico de pesquisa em questão. Essa medida pode auxiliar na definição e maturidade do protocolo da revisão e também na experiência inicial do pesquisador. Durante a execução da busca, um dos motivos para utilizar a literatura cinzenta é a diminuição do viés de publicação, como mencionado na Questão 5, que se refere ao problema de que resultados positivos são mais publicados do que resultados negativos. Ressalta-se que o conceito de resultados positivos e negativos pode depender do ponto de vista do pesquisador. Mesmo assim, viés de publicação pode surgir do modo como um determinado estudo primário é avaliado [10], ou seja, o crivo utilizado na avaliação dos resultados impacta diretamente as hipóteses e conclusões criadas a partir do estudo primário conduzido.

Uma das maneiras de se realizar a busca na literatura cinzenta é utilizar o motor de busca *Google Scholar*¹⁰, disponível a partir de novembro de 2004, que realiza buscas multidisciplinares por trabalhos revisados, dissertações de mestrado, teses de doutorado, livros, artigos, resumos e outros, a partir de publicadores acadêmicos, sociedades profissionais, repositórios de pré-publicações, universidades e outras organizações escolares [65]. Há também outras maneiras, como o *arXiv*¹¹, que indexa publicações cujos temas são diversos (por exemplo, Ciência da Computação e Estatística).

Sobre a literatura cinzenta, pode-se concluir que ela serve como complemento de uma estratégia de busca. Assim, seu uso pode auxiliar antes, durante e depois do processo de pesquisa, a fim de fornecer uma visão geral da pesquisa, diminuição do viés de publicação e comparação de resultados obtidos com os disponíveis nessa literatura. Como os estudos primários recuperados da literatura cinzenta complementam os resultados dos publicadores comerciais, esses estudos também devem ser avaliados mesmo que sejam preliminares [64], ou até mesmo contenham baixa qualidade em termos de nível de avaliação [4].

Questão 9: O que é bola de neve (*snowballing*)?

¹⁰ <https://scholar.google.com.br>

¹¹ <http://arxiv.org>

Snowballing, assim como a busca automática e a busca manual, é empregada para identificar estudos primários e, para isso, utiliza-se a lista de citações e referências dos estudos primários já selecionados. No trabalho de Wohlin [53], o autor apresenta diretrizes para a execução de buscas utilizando a *snowballing*. O autor discute e formaliza a estratégia e sua execução adequada dividindo-a em duas etapas, *snowballing* reverso e *snowballing* avante, sendo ambas complementares.

Snowballing reverso consiste em avaliar a lista de referências de um estudo primário relevante procurando por outros estudos primários relevantes.

Snowballing avante consiste em avaliar a lista de citações para um estudo primário já considerado relevante. É necessário utilizar recurso computacional para identificar a lista de citações, sendo que motores de busca e bases bibliográficas fornecem esse recurso. Um grande desafio é como avaliar o quanto completa está a lista encontrada. O número de trabalhos indexados nas bases bibliográficas é limitado e isso pode comprometer a completude da lista. Kitchenham e Brereton [63] utilizaram *snowballing* avante nos estudos primários mais relevantes de suas RSs e afirmam que esse é um meio efetivo para identificar novos estudos primários.

Segundo Petersen et al. [49], o pesquisador pode obter melhores resultados em suas buscas ao combinar buscas utilizando a *string* de busca com estratégias secundárias de busca como *snowballing*. As buscas automáticas são estratégias eficazes para a obtenção de estudos primários em uma RS, mas existem limitações para seus resultados. As limitações estão relacionadas às *strings* e incluem sensibilidade aos sinônimos de termos e aos idiomas. A qualidade dos repositórios nos quais as buscas são executadas também influencia nos resultados de uma busca utilizando-se *strings*. *Snowballing* é uma estratégia que pode ser utilizada para reduzir esses problemas, uma vez que é independente de palavras-chave e de bases bibliográficas.

O pesquisador pode adotar duas abordagens para aplicar *snowballing*: utilizá-la como estratégia primária de pesquisa, partindo de um conjunto inicial prévio, como proposto por Wohlin [53]; ou como estratégia secundária, para melhoria de resultados de busca já executada com a *string* de busca. Para executar *snowballing* como uma estratégia primária de pesquisa, o pesquisador deverá definir um conjunto inicial de estudos primários [66]. Esse conjunto deverá conter uma seleção inicial de estudos primários relevantes, nos quais se aplicam *snowballing* reverso e avante. Novos estudos primários serão encontrados e para cada um dos novos estudos primários, o pesquisador deverá repetir o processo até não encontrar

mais estudos relevantes. Em uma estratégia primária de pesquisa, a escolha do conjunto inicial é de extrema importância para atingir resultados adequados [53]. O conjunto inicial deve ser heterogêneo, caso contrário, uma grande quantidade de estudos primários relevantes poderá ser ignorada. Um conjunto heterogêneo, nesse caso, deve ser entendido como um conjunto de publicações pertencentes a diferentes grupos de pesquisa.

Para executar *snowballing* como estratégia secundária, inicialmente o pesquisador deverá aplicar a busca automática utilizando a *string* de busca. Para cada estudo primário, deve então aplicar *snowballing* para localizar novos estudos primários. O processo repete-se até que não sejam mais encontrados estudos relevantes. Como um exemplo, *snowballing* pode ser utilizada para validar os resultados obtidos com as buscas com a *string*. Quando o pesquisador aplica *snowballing* em algum estudo primário previamente obtido pela busca com *string* e novos estudos relevantes são encontrados, a seguinte questão deve ser respondida: *Por qual motivo os novos estudos encontrados não estão contidos nos resultados obtidos com a busca inicial?* Se um estudo primário relevante for localizado após a aplicação de *snowballing* e estiver indexado nas bases bibliográficas previamente consultadas, isso significa que existem palavras-chave não contidas na *string* utilizada para busca. Se o estudo estiver indexado em uma dada base na qual a busca com *string* não foi executada, isso é indicativo de que essa base potencialmente relevante pode ter sido esquecida. Em ambos os casos, o pesquisador deverá tomar as providências cabíveis e executar novamente as buscas para garantir melhores resultados. Se os novos estudos primários encontrados por *snowballing* não estiverem indexados em repositórios, o pesquisador poderá incluí-los em sua pesquisa. Para esses casos, não é necessário atualizar a *string* de busca ou refazer as buscas.

Caso executada de forma inadequada, *snowballing* pode induzir a viés. Sendo assim, é necessário aplicá-la com cautela. *Snowballing* pode recuperar estudos que não estão indexados nas principais fontes e cabe ao pesquisador avaliar os estudos primários encontrados de forma criteriosa, principalmente com relação à qualidade das evidências. Segundo Wohlin [53], a qualidade dos estudos encontrados com *snowballing* deve ser rigorosamente avaliada antes que os mesmos sejam inclusos na RS. Um exemplo é a grande quantidade de literatura cinzenta que pode ser encontrada, cabendo ao pesquisador definir se esses estudos podem ou não compor a RS. Detalhes sobre o impacto do uso de literatura cinzenta na pesquisa são discutidos na Questão 8.

Questão 10: Por que é necessário documentar o processo de busca?

Assim como as demais atividades do processo de execução de RS, a atividade de busca deve ser documentada apropriadamente para garantir transparência e permitir que os leitores possam verificar a sua completude e abrangência [6] [10]. Não somente os resultados finais devem ser registrados, mas também os resultados intermediários como, estudos primários que não foram selecionados para compor o conjunto final e a aplicação e interpretação dos critérios de exclusão nesses estudos.

A documentação do processo de busca viabiliza dois requisitos fundamentais de uma RS [6] [10] [67]: auditabilidade e reproduzibilidade. Espera-se que tanto o processo quanto os resultados possam ser auditados para verificação da exatidão e imparcialidade na seleção de estudos e síntese de resultados [10]. Além disso, espera-se que uma RS possa ser reproduzida por outros (ou até os mesmos) pesquisadores, com o intuito de comparar se os resultados corroboram os iniciais ou de estender a RS original com novas evidências. Claramente, a esperada similaridade de resultados de ambas as revisões é fortemente dependente da objetividade observada na descrição dos critérios de inclusão e exclusão de estudos, na definição adequada da estratégia de busca e na clareza dos procedimentos de extração e síntese de dados.

De acordo com Kitchenham e Charters [7], os seguintes itens de informação devem compor a documentação do processo de busca: as fontes de informação pesquisadas, que incluem desde as bases bibliográficas quanto os nomes dos especialistas consultados; a *string* de busca e suas adaptações para bases bibliográficas específicas; a data em que a busca foi realizada; o período considerado (por exemplo, ano de publicação dos estudos primários); e informações básicas de cada estudo primário recuperado com a busca (por exemplo, título, veículo de publicação, autores e palavras-chave). Esses itens de informação serão fundamentais para que, a qualquer momento, revisões e refinamentos do processo de busca e de etapas posteriores da RS sejam realizados. Para armazenar tais informações, os pesquisadores podem utilizar planilhas eletrônicas convencionais, ferramentas de gerenciamento de referências (por exemplo, *EndNote*¹² e *Mendeley*¹³) ou ferramentas específicas para apoiar o processo de RS

¹² <http://www.endnote.com>

¹³ <http://www.mendeley.com>

como, por exemplo, a *StArt*¹⁴.

A documentação do processo de busca também ajuda a registrar lições aprendidas como, por exemplo, o uso efetivo de bases bibliográficas e a definição dos termos que compõem a *string* de busca. Um fato comumente observado ao se utilizar motores de busca (vide Tabela 3.1) é a sobreposição de resultados por eles gerados, pois os motores de busca recuperam estudos fornecidos por variadas bases indexadas. Outro fato recorrente refere-se à inclusão de siglas na *string* de busca, sendo que algumas siglas podem referenciar-se a conceitos diferentes, causando substancial aumento na quantidade de estudos a serem analisados na seleção. Como exemplo, cita-se a decisão de Oliveira et al. [45] de não incluir a sigla *SOC*, que para o tópico pesquisado pelos autores, significa *Service-Oriented Computing*. Porém, a mesma sigla é também empregada para os conceitos *State of Charge*, *Stochastic Optimal Control*, *System-on-chip* e *Society*.

Considerações finais

A busca por estudos primários deve ser eficaz e eficiente para que, a um custo viável, identifiquem-se estudos que de fato contribuam para atingir os objetivos definidos para a RS. O foco deste capítulo foi na apresentação de diversas questões relacionadas à definição de uma estratégia abrangente de busca, desde a criação da *string* de busca até a documentação da estratégia busca adotada. Em resumo, os principais pontos de aprendizagem deste capítulo foram:

- Uma estratégia abrangente de busca deve considerar variadas fontes de estudos primários, desde bases bibliográficas até literatura cinzenta;
- A busca piloto é essencial para refinar e/ou estender a *string* de busca e aumentar a chance de que ela viabilize uma busca eficaz e eficiente;
- As buscas automática, manual e *snowballing* são complementares;
- A consulta a especialistas pode ser útil tanto para a identificação de estudos primários relevantes quanto para a avaliação desses estudos; e
- A documentação do processo de busca é imprescindível para garantir auditabilidade, transparência e reproduzibilidade da RS.

¹⁴ http://lapes.dc.ufscar.br/tools/start_tool

4. Seleção e avaliação de estudos

Katia Romero Felizardo
Érica Ferreira de Souza
Ricardo de Almeida Falbo

Uma busca abrangente pode resultar em um grande número de estudos candidatos (visite o Capítulo 3 para mais detalhes sobre estratégias de busca). Desses estudos, devem ser identificados quais são relevantes e quais devem ser desconsiderados. Para isso, aplicam-se critérios de seleção (critérios de inclusão e exclusão) e de qualidade. Todos os critérios são estabelecidos antes de se iniciar o processo de RS e são descritos no protocolo (veja Capítulo 2 para esclarecimentos adicionais sobre o protocolo). Os critérios de seleção especificam as características que os estudos devem conter para serem incluídos ou excluídos. O objetivo desses critérios é assegurar que estudos relevantes para responder às questões de pesquisa da RS sejam incluídos e que nenhum estudo importante seja excluído. Já os critérios de qualidade objetivam avaliar aspectos metodológicos dos estudos.

O processo de seleção é composto por duas etapas: a seleção inicial e a seleção final. Inicialmente, os critérios de seleção são aplicados em todos os estudos candidatos após a leitura de seu título e resumo. Nessa etapa, os estudos só serão excluídos se eles atenderem claramente um ou mais dos critérios de exclusão. Estudos incluídos na primeira etapa são examinados com mais detalhes na segunda etapa, que consiste na aplicação dos critérios de seleção após a leitura do texto completo dos estudos. É recomendado que, após a seleção final, a qualidade dos estudos incluídos seja avaliada com base nos critérios de qualidade. Assim, aspectos como o tipo de pesquisa realizada (*survey*, estudo de caso, experimento, etc) e o método para coleta de dados adotado (histórico, de observação, controlado, etc) pelo estudo, dentre outros, são avaliados. Neste capítulo é descrito como selecionar estudos relevantes para responder às questões de pesquisa de uma RS. É também destacada a necessidade de realizar a avaliação da qualidade dos estudos.

Questão 1: O que são critérios de seleção?

Uma das características que difere a RS de outros métodos de revisão é que a RS segue um método rigoroso de seleção de estudos, com base nos critérios de seleção, que abrangem tanto critérios de inclusão quanto critérios de exclusão. A qualidade dos resultados da RS depende dos estudos incluídos. Portanto, um passo crítico é a definição dos critérios de seleção.

Os critérios de inclusão definem características que um estudo deve conter para ser considerado relevante para aquela RS. Por outro lado, os critérios de exclusão estabelecem características para excluírem estudos considerados irrelevantes no contexto definido. Esses critérios são criados com base nas questões de pesquisa da RS e devem ser descritos previamente no protocolo e rigorosamente aplicados. Além dos critérios, também deve estar descrito no protocolo como esses critérios serão aplicados, por exemplo, quantos revisores irão participar de cada etapa da seleção ou como os desacordos entre os revisores serão resolvidos.

Com relação aos critérios de inclusão, eles devem ser definidos de forma alinhada com o objetivo da RS. Pode-se definir um único critério de inclusão ou vários. Quando mais de um critério de inclusão for definido, deve ficar claro se os critérios de inclusão são alternativos (ou seja, se um estudo satisfizer um dos critérios, então ele deverá ser selecionado) ou se os critérios deverão ser aplicados em conjunto (ou seja, para um estudo ser incluído, ele deverá satisfazer todos os critérios de inclusão).

Com relação aos critérios de exclusão, considera-se que um critério de exclusão direto é o não atendimento do(s) critério(s) de inclusão (ou o não atendimento de nenhum dos critérios de inclusão, no caso de haver vários critérios de inclusão alternativos). Pode ser definido também um conjunto de critérios de exclusão gerais que vem sendo adotado em RSs publicadas na literatura, tais como:

- **O estudo não possui um resumo:** Alguns estudos não têm um resumo, o que impossibilita que uma seleção inicial seja feita sem que se tenha que ler o texto completo. Em casos como esse, usando esse critério de exclusão, um estudo que não tenha um resumo será excluído;
- **O estudo é publicado apenas como resumo:** Muitas vezes, um estudo é publicado apenas na forma de resumo. Isso acontece, por exemplo, em alguns casos em que o estudo é publicado como um pôster em um evento. É importante não confundir resumos com artigos curtos (*short papers*). Um artigo curto tem resumo e também um corpo, ainda que o corpo possa ser pequeno (geralmente duas ou três páginas). Um estudo publicado apenas como resumo, por outro lado, não tem corpo. Todo o seu conteúdo restringe-se ao resumo;
- **O estudo não está escrito em inglês:** Uma vez que o inglês é a língua atualmente usada para a comunicação científica, considerar estudos escritos em outras línguas pode comprometer a reproduzibilidade ou mesmo o uso dos resultados de uma RS por

outros pesquisadores. Assim, uma boa opção é excluir tais estudos. Vale ressaltar que alguns estudos têm título, resumo e palavras-chave escritos em inglês, mas o seu corpo está escrito em outra língua. Nesse caso, mesmo que um estudo tenha sido selecionado na seleção inicial (aquele que considera apenas título e resumo), ele deverá ser excluído na seleção final;

- **O estudo é uma versão mais antiga de outro estudo já considerado:** É importante não considerar evidências repetidas de um mesmo estudo, publicado em diferentes veículos de divulgação, porque ele pode criar distorções nas conclusões da RS [10] [45]. Algumas vezes, um mesmo estudo é publicado em mais de um veículo de publicação. Isso acontece, por exemplo, quando um estudo é publicado em um evento e, posteriormente, uma versão estendida é publicada em um periódico. Nesses casos, só deve ser considerada a publicação mais recente. Vale observar, contudo, que um estudo só deve ser eliminado por esse critério se o estudo mais recente considerado tratar todo o conteúdo tratado na versão mais antiga. Estudos com alguma interseção, mas sem que o conteúdo da versão mais antiga esteja completamente tratado na versão mais recente, não devem ser excluídos;
- **O estudo não é um estudo primário:** Uma RS visa coletar evidências de estudos primários. Assim, estudos que não são resultado de pesquisa não devem ser considerados. Enquadram-se nesse caso editoriais, resumos de palestras, sumário de anais de eventos científicos (conferências, *workshops*, simpósios, etc), tutoriais e outros estudos secundários (outras RSs, mapeamentos sistemáticos da literatura ou mesmo estudos apresentando revisões informais da literatura); e
- **Não foi possível ter acesso ao estudo:** A maior parte das bases bibliográficas tem acesso controlado. Assim, é possível que o revisor não tenha acesso a alguns dos estudos selecionados. De maneira geral, informações de título, resumo e palavras-chave estão sempre disponíveis, mas o mesmo não ocorre com o texto completo. Por exemplo, pesquisadores brasileiros com acesso apenas ao Portal de Periódicos da CAPES¹⁵ não tem acesso aos estudos publicados na *Springer Link*¹⁶. Quando o revisor não tiver acesso ao texto completo, esse critério pode ser usado para excluir o estudo da seleção final. Vale ressaltar, contudo, que esse critério deve ser aplicado como último recurso. Antes de excluir o estudo, é aconselhável que o revisor use outros meios para tentar ter acesso ao estudo, tais como solicitar diretamente aos autores o

¹⁵ <http://www.periodicos.capes.gov.br>

¹⁶ <http://www.springeropen.com/search>

envio do estudo, buscar em redes sociais acadêmicas voltadas para pesquisadores, tais como *Research Gate*¹⁷ e *Academia*¹⁸, ou procurar encontrar o artigo na web usando mecanismos de busca gerais como o *Google*. O trabalho de Cruz et al. [68] é um exemplo de RS que eliminou estudos por falta de acesso aos textos completos.

É importante observar que definir critérios de exclusão relacionados a um período de tempo pode limitar o número de estudos incluídos, principalmente se o prazo estabelecido for estreito. No entanto, existem situações nas quais existe uma razão para essa escolha, por exemplo, apresentar um panorama dos últimos 10 anos sobre um dado tópico de interesse, como a RS de Conry et al. [69].

Os critérios de seleção devem ser avaliados por meio de um teste piloto. Nesse caso, os critérios de inclusão e exclusão são aplicados em um número reduzido de estudos candidatos, tipicamente considerando uma única base (por exemplo, *IEEE Xplore*). Um dos objetivos é refinar os critérios, assim como treinar os revisores nos procedimentos de seleção e na aplicação dos critérios definidos [18]. Mesmo após o teste piloto, os critérios de seleção podem mudar ao longo da atividade de seleção, mas, se alterados, devem ser atualizados no protocolo e aplicados de forma retroativa nos estudos avaliados antes da modificação. Revisores podem, inclusive, adotar critérios de inclusão e exclusão gerais para a primeira etapa do processo de seleção e critérios mais específicos para a segunda etapa. Por exemplo, Bano e Zowghi [70] adotaram diferentes critérios para a seleção inicial e para a seleção final de estudos.

Os critérios de seleção podem ser amplos ou mais específicos. Se, por exemplo, os critérios de inclusão forem muito gerais, muitos estudos (inclusive estudos de menor qualidade) poderão ser incluídos e, consequentemente, a confiança no resultado final é diminuída. Por outro lado, se esses critérios forem muito rigorosos, os resultados podem ser baseados em um número pequeno de estudos, o que dificulta a generalização dos achados [71] em situações em que o pesquisador enfrenta uma escassez de estudos relevantes para responder às questões de pesquisa. Por exemplo, há poucos estudos experimentais sobre o uso de mineração visual de texto (*Visual Text Mining - VTM*) para apoiar a seleção de estudos no processo de RS [72]. Assim, se o objetivo da RS é avaliar os efeitos do uso de VTM, os critérios de inclusão poderiam ser mais amplos para incluir outros tipos de pesquisa, como estudos de caso ou quasi-experimentos.

¹⁷ <http://www.researchgate.net>

¹⁸ <http://www.academia.edu>

Para ilustrar a elaboração de critérios de seleção de estudos, considere o trabalho de Oliveira et al. [45]. Nessa RS, o objetivo foi identificar estudos que relatam o desenvolvimento de sistemas robóticos orientados a serviços, ou seja, sistemas robóticos que foram projetados com base em SOA. Para selecionar estudos em concordância com esse objetivo, foram definidos os seguintes critérios de inclusão alternativos (ou seja, basta um deles ser satisfeito para que um estudo seja incluído):

1. O estudo propõe ou relata o projeto e/ou desenvolvimento de um sistema robótico orientado a serviços; ou
2. O estudo propõe ou relata uma nova tecnologia para o desenvolvimento de sistemas robóticos orientados a serviços; ou
3. O estudo propõe ou relata um processo, método, técnica, arquitetura de referência ou diretrizes de ES para apoiar o projeto e/ou desenvolvimento de sistemas robóticos orientados a serviços.

Os critérios de exclusão de Oliveira et al. [45] foram:

1. O estudo relata o desenvolvimento de um sistema robótico sem o uso de SOA;
2. O estudo apresenta contribuições em outras áreas que não a robótica;
3. O estudo é uma versão anterior de um estudo mais completo sobre a mesma pesquisa;
4. O estudo é a descrição de um curso, editorial, resumo de palestra, *workshop* ou tutorial.

Como é possível notar, os dois primeiros critérios de exclusão são relativos ao não atendimento de nenhum dos critérios de inclusão.

Questão 2: Quais são as etapas da seleção de estudos primários?

A Figura 4.1 mostra as etapas da seleção, que é iniciada após a identificação dos estudos que potencialmente podem ser incluídos na revisão. Esses estudos são retornados tanto das buscas em bases bibliográficas quanto de outras fontes, por exemplo, estudos buscados manualmente. É importante ressaltar que um mesmo estudo pode estar indexado em mais de uma base, de modo que existirão estudos duplicados. Assim, antes de iniciar efetivamente a seleção, a duplicidade de estudos deve ser eliminada. A seleção é geralmente realizada em duas etapas: seleção inicial e seleção final.

Durante a seleção inicial, os critérios de inclusão e exclusão são aplicados em todos os estudos identificados, por meio da avaliação de seu título e resumo. Muitos pesquisadores que

realizam RSs relatam que têm encontrado dificuldades para identificar se um dado estudo é ou não relevante apenas com a leitura do título e resumo [9]. Realmente, resumos com ausência de informações, tais como o método de pesquisa adotado ou os resultados alcançados, podem dificultar a seleção [9] [14] [59] [73] [74]. Assim, na dúvida sobre a inclusão ou não de um estudo durante a primeira etapa, deve-se optar pela sua inclusão. Já durante a seleção final, os critérios de inclusão e exclusão são aplicados nos estudos incluídos na primeira etapa, por meio da avaliação de seu texto completo.

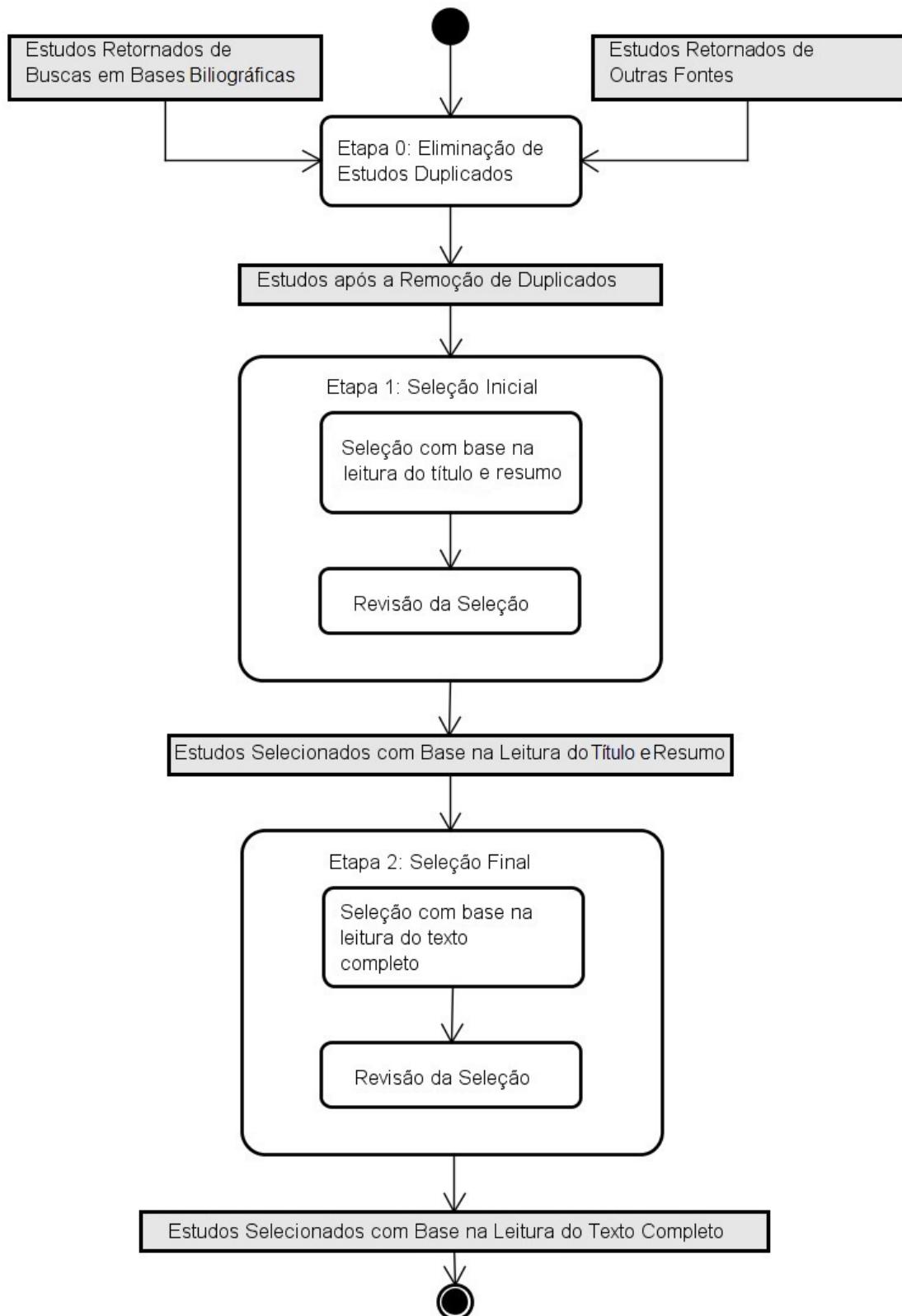


Figura 4.1: Etapas da seleção de estudos.

Após cada uma das etapas de seleção, seja a inicial ou a final, deve-se realizar uma revisão da seleção. Essa atividade é indicada para aumentar a confiabilidade e transparência da seleção e visa evitar que estudos relevantes sejam excluídos. Imagine uma seleção sendo realizada por

dois ou mais pesquisadores de forma independente. Nesse caso, um mesmo estudo pode ter sido incluído por um e excluído pelo outro. Durante a atividade de revisão da seleção, esses estudos serão reavaliados pelos próprios pesquisadores, ou até mesmo com a ajuda de um pesquisador externo. Um exemplo é a revisão de Galster et al. [42], na qual os estudos foram analisados por dois pesquisadores e as diferenças solucionadas de forma colaborativa. No caso de pesquisadores que conduzem a seleção de forma isolada, a revisão da seleção pode ser usada para que o pesquisador possa discutir suas decisões com outro(s) pesquisador(es). Alternativamente, o próprio pesquisador pode reavaliar uma amostra aleatória de estudos para determinar a consistência das decisões. Essa amostra aleatória deve ser formada por pelo menos 1/3 dos estudos. No entanto, é importante que os pesquisadores marquem explicitamente aqueles estudos que são objetos de dúvida, os quais devem ser obrigatoriamente revisados. Se a quantidade de estudos retornados não for muito grande e houver vários pesquisadores envolvidos na seleção, mais de 1/3 dos estudos, e até mesmo todos os estudos retornados, podem ser reavaliados. Um exemplo é a revisão realizada por Hall et al. [75], na qual a seleção foi avaliada usando um conjunto aleatório de estudos. Por fim, após a seleção final, é recomendado que a qualidade dos estudos selecionados seja avaliada. Essa avaliação baseia-se em critérios de qualidade, descritos na Questão 3.

Fabbri et al. [18] destacam que durante a seleção, palavras-chave ou suas variações que não foram adotadas inicialmente nas buscas podem ser identificadas. Isso pode acontecer por diferentes razões, como o revisor ser novato na área de pesquisa ou a palavra-chave ou suas variações ser um novo termo dessa área. Se novas palavras-chave ou termos forem descobertos, eles devem ser considerados e desencadearão uma reexecução de atividades, a começar pela busca por estudos. Consequentemente, a seleção também deverá ser reexecutada para os novos estudos encontrados. Também é possível que novos campos para o formulário de extração de dados sejam identificados durante a leitura dos estudos. Nesse caso, o formulário, elaborado inicialmente durante a criação do protocolo, deve ser também atualizado.

Na RS conduzida por Oliveira et al. [45], a seleção de estudos foi conduzida em duas etapas: a seleção inicial e a seleção final. Inicialmente, os critérios de inclusão e exclusão foram aplicados após a leitura do título e resumo de todos os estudos retornados na busca. Como resultado, estudos que atendiam aos critérios de inclusão e que possivelmente responderiam às questões de pesquisa foram considerados relevantes e incluídos. Por exemplo, o estudo de Blake et al. [76] foi incluído na primeira etapa devido ao seguinte critério de inclusão: *O estudo propõe ou relata uma nova tecnologia para o desenvolvimento de sistemas robóticos*

orientados a serviços. De fato, esse estudo propõe um mecanismo para interligar serviços web tradicionais aos componentes desenvolvidos para o ambiente *Robot Operating System* (ROS), possibilitando a criação de sistemas robóticos orientados a serviço integráveis à Internet. Já estudos que atendiam aos critérios de exclusão foram eliminados. Nesse contexto, o estudo de Doran e Gokhale [77] foi excluído com base no seguinte critério de exclusão: *O estudo apresenta contribuições em outras áreas que não a robótica*. A justificativa para essa exclusão foi de que o estudo estava relacionado com à temática *Internet Bot*, que é um programa de computador desenvolvido para coletar informações da Internet para prever acontecimentos futuros. Durante a segunda etapa, cada estudo incluído na etapa inicial foi lido na íntegra. Como exemplo, tem-se o estudo de Tsai et al. [78], que foi incluído com base em dois critérios de inclusão: *O estudo propõe ou relata o projeto e/ou desenvolvimento de um sistema robótico orientado a serviços*; e *O estudo propõe ou relata uma nova tecnologia para o desenvolvimento de sistemas robóticos orientados a serviços*. Isso porque o estudo elaborou um *framework* baseado no ambiente *Microsoft Robotics Studio* para o desenvolvimento de sistemas robóticos orientados a serviços e também detalhava o desenvolvimento de um sistema robótico baseado em SOA como forma de ilustrar o *framework* proposto. Por outro lado, o estudo de Zhang et al. [79] foi excluído devido ao critério de exclusão: *O estudo relata o desenvolvimento de um sistema robótico, sem uso de SOA*. No caso específico desse estudo, o trabalho desenvolvido não estava relacionado à SOA, mas ao conceito de robótica de serviço. Na dúvida sobre a inclusão ou exclusão desse estudo, o mesmo foi analisado por dois pesquisadores de forma independente. A decisão final foi tomada por consenso.

Questão 3: O que é a avaliação da qualidade de estudos primários?

A qualidade de uma RS depende da qualidade dos estudos primários incluídos. A qualidade de cada estudo está diretamente associada à sua qualidade metodológica, que pode ser medida por aspectos como: validade interna (medir efetivamente o que se deseja mensurar), validade externa (poder de generalizar os resultados), relevância do tema de pesquisa (questões de pesquisa e objetivos bem definidos e embasados na literatura), adoção de métodos que conduzam aos objetivos propostos, entre outros.

Após a seleção final (aplicação dos critérios de inclusão e exclusão após a leitura do texto completo dos estudos), muitas vezes, os revisores realizam a avaliação da qualidade desses estudos. Staples e Niazi [47] dizem que após a seleção final, uma avaliação da qualidade permite aos pesquisadores analisem mais detalhadamente os estudos.

A avaliação da qualidade é realizada por meio de formulários desenvolvidos pelos próprios pesquisadores. Esses formulários são compostos pelos critérios de qualidade, ou seja, perguntas sobre aspectos metodológicos de cada estudo [10] [47]. Embora exemplos de formulários de qualidade possam ser encontrados na literatura [7] [10] [17] [63], recomendamos que os mesmos sejam elaborados especificamente para cada revisão, de acordo com o tema a ser abordado e em conformidade com as questões de pesquisa.

Um exemplo de RS que fez uso de critérios de qualidade foi a conduzida por Mendes et al. [43], cujo objetivo foi identificar o melhor modelo de estimativa de esforço no desenvolvimento de software web dentre dois modelos: modelos construídos usando conjuntos de dados de várias organizações; e modelos construídos a partir de dados de organizações individuais. O formulário construído por Mendes et al. [43] é composto por questões que avaliam aspectos metodológicos dos estudos, tais como: *Os métodos utilizados para construção dos modelos de estimativas foram totalmente definidos (ferramentas e métodos utilizados)? O resultado da análise de sensibilidade (ou análise de resíduos) foi usado para remover pontos de dados anormais, se necessário?* Analisando as questões do formulário de Mendes et al., é possível observar que o grande desafio da avaliação da qualidade é a elaboração de critérios objetivos, que garantam que a RS se mantenha reproduzível. Por isso, apesar de recomendadas, avaliações de qualidade nem sempre são realizadas. Por exemplo, na RS conduzida por Oliveira et al. [45], essa atividade não foi considerada. Isso se deve ao fato desse tipo de avaliação poder introduzir viés na RS, pois critérios de qualidade, na maioria das vezes, são subjetivos e difíceis de serem aplicados por pesquisadores inexperientes.

Se são atribuídas escalas numéricas para os itens que compõem os formulários, então valores numéricos da qualidade (índices de qualidade) podem ser obtidos. Por exemplo, se questões de um formulário puderem ser respondidas por *sim* ou *não* (correspondendo a uma pontuação de 1 e 0, respectivamente), então, para um questionário composto por quatro questões, o nível de qualidade de um estudo variaria de 0 (estudo de *baixa qualidade*) a 4 (estudo de *alta qualidade*). Esses índices podem ser usados para excluir estudos. No entanto, pode valer mais a pena utilizá-los para avaliar o peso que cada estudo possui dentro do contexto geral dos estudos incluídos. Ou seja, com base nos índices de qualidade, encorajamos que os pesquisadores discutam sobre a qualidade (ou maturidade) dos estudos incluídos e o impacto de cada um deles nos resultados gerais da RS. Um estudo pode ser classificado em diferentes níveis de maturidade, como discutido por Wieringa et al. [80]. Nesse sentido, a avaliação da qualidade também pode ser usada para auxiliar a análise e a síntese dos dados. Nesse caso, os índices de qualidade são usados para identificar subgrupos de estudos. A diferença de

qualidade entre os subgrupos é investigada para verificar se essa diferença pode explicar resultados divergentes identificados nos estudos incluídos.

Questão 4: Quantos pesquisadores devem participar da seleção de estudos primários?

Decidir pelos estudos a serem incluídos em uma RS envolve julgamento por parte do(s) pesquisador(es)/revisor(es). Além disso, um único equívoco na aplicação dos critérios de seleção é suficiente para um estudo relevante ser excluído de uma RS, o que pode comprometer a qualidade dos resultados obtidos. Assim, é muito importante que mais de um revisor participe da seleção. Em especial, quando a RS estiver sendo conduzida por um pesquisador não especialista no tema de estudo, é fundamental que tanto a seleção quanto a avaliação da qualidade dos estudos sejam conduzidas por mais de um pesquisador, sendo que algum deles deve ser especialista no tema.

Na prática, a estratégia de seleção pode variar de revisão para revisão, dependendo da experiência dos revisores tanto no tema da RS quanto no próprio processo de RS. No entanto, se a seleção é realizada por pelo menos dois revisores, há uma redução na possibilidade de que estudos relevantes sejam descartados [81].

A seleção pode ser realizada de diferentes formas, descritas a seguir.

- **Configuração 1:** dois ou mais pesquisadores de forma independente realizam a seleção e compararam os resultados. Conflitos são solucionados por consenso. Por exemplo, na revisão conduzida por Dieste e Juristo [82], dois pesquisadores checaram de maneira independente os estudos;
- **Configuração 2:** o total de estudos é avaliado por dois (ou mais) pesquisadores, que realizam a seleção de forma independente. Um pesquisador externo avalia o resultado da seleção e classifica os estudos que tiveram divergências quanto à seleção feita pelos pesquisadores. Como já mencionado na Questão 2, sugere-se que o conjunto de estudos a serem reavaliados pelo pesquisador externo seja composto de pelo menos 1/3 dos estudos, incluindo necessariamente os estudos com classificações divergentes. Para completar esse conjunto, os demais estudos são escolhidos aleatoriamente. Pode-se inclusive optar pela revisão da totalidade dos estudos, caso esse número não seja muito grande e haja disponibilidade de recursos (tempo e revisores). Em caso de RS sendo conduzida por alunos de mestrado ou doutorado, é muito comum que os orientadores desempenhem o papel de avaliador, mas outros pesquisadores com conhecimento sobre o tema da RS podem atuar como pesquisadores externos. Enfim, a condição básica para atuar como avaliador da seleção é ser um especialista no tema

da revisão. Por exemplo, durante a seleção final da RS de Cruz et al. [68], dois pesquisadores trabalharam independentemente. Um pesquisador externo, que não participou da seleção inicial auxiliou na resolução das divergências. Em casos nos quais um acordo não foi possível, as diferenças foram solucionadas em uma reunião de consenso;

- **Configuração 3:** dois pesquisadores em conjunto realizam a atividade de seleção, de forma semelhante a uma sessão de programação em pares, adotada nos métodos de desenvolvimento ágeis;
- **Configuração 4:** um pesquisador realiza a atividade de seleção e um segundo pesquisador (ou vários pesquisadores) avalia o resultado. Nesse caso, deve ser definida uma amostra de estudos a serem avaliados (contendo pelo menos 1/3 dos estudos), incluindo aqueles estudos em que o pesquisador responsável pela seleção teve dúvidas quanto à sua seleção. Conflitos devem ser solucionados por consenso. Por exemplo, na revisão conduzida por MacDonell e Shepperd [39], um pesquisador executou a atividade de seleção e um segundo pesquisador revisou a decisão de incluir ou não um subconjunto dos estudos.

Para resolver divergências sobre a inclusão de um estudo, é aconselhável obter informações adicionais sobre o trabalho. Adicionalmente, as razões que geraram discordâncias devem ser exploradas, pois elas podem revelar a necessidade de, por exemplo, melhorar a descrição dos procedimentos de seleção, assim como refinar os critérios de inclusão e exclusão. Na revisão realizada por Hall et al. [75], os critérios de inclusão e exclusão foram melhorados após as causas das discordâncias entre os revisores terem sido analisadas.

A confiabilidade da seleção pode ser medida pelo teste Kappa, que aponta a confiabilidade entre classificações feitas por diferentes pesquisadores. Esse teste considera que concordância por acaso pode ocorrer na seleção (estudos incluídos e excluídos) feitas por dois ou mais pesquisadores [83]. Por exemplo, seja uma RS contendo 10 estudos a serem avaliados por dois pesquisadores (pesquisador 1 e pesquisador 2), sendo que cada pesquisador pode classificar os estudos como incluído ou excluído. A Tabela 4.1 sumariza um possível resultado da seleção.

Aplicando-se o teste Kappa, obtém-se um coeficiente de mesmo nome, o qual é calculado pela seguinte expressão matemática:

$$(4.1) \quad Kappa = \frac{Co - Ca}{1 - Ca}$$

onde:

- Co é a concordância observada (Co) ou a taxa de concordância real;
- Ca é a concordância ao acaso, ou aquela esperada se a classificação dos estudos como incluído ou excluído fosse feita de maneira aleatória.

Tabela 4.1: Resultado da seleção contendo estudos incluídos e excluídos.

	Pesquisador 1	Pesquisador 2
Estudo 1	I	E
Estudo 2	I	I
Estudo 3	I	I
Estudo 4	I	E
Estudo 5	I	I
Estudo 6	I	I
Estudo 7	E	E
Estudo 8	E	E
Estudo 9	E	I
Estudo 10	E	E
Legenda: I = Incluído; E = Excluído		

O cálculo do Co e Ca para o exemplo da Tabela 4.1 são mostrados na Figura 4.2. É possível observar que os dois pesquisadores tiveram uma concordância observada (Co) de 70% (ou seja, 40% + 30%) (conforme ilustrado na Figura 4.2(a)). Pode-se notar que em 40% das vezes, os dois pesquisadores classificaram os mesmos estudos como incluídos (observe na Tabela 4.1 que quatro estudos foram classificados como incluídos por ambos os pesquisadores – Estudos 2, 3, 5 e 6), e em 30% das vezes, os dois classificaram os mesmos estudos como excluídos (observe na Tabela 4.1 que três estudos foram classificados como excluídos por ambos os pesquisadores – Estudos 7, 8 e 10).

O cálculo da concordância ao acaso (Ca) é baseado nos marginais totais. O marginal total de uma linha é calculado pela soma dos valores das células dessa linha e armazenado no final dessa linha (célula com fundo cinza na Figura 4.2(b)). De maneira similar, o marginal total de uma coluna é calculado pela soma dos valores das células dessa coluna e armazenado no final da mesma (célula com fundo cinza na Figura 4.2(b)). A seguir, calcula-se a concordância ao acaso dos dois pesquisadores terem incluído os mesmos estudos, ou seja, $Ca(I)$ (conforme ilustrado na Figura 4.2(c)). Para isso, multiplicam-se os marginais totais da linha (I) e da coluna (I). No nosso exemplo, $Ca(I)$ é 30% (ou seja, 60% * 50%). Na mesma linha, calcula-se a concordância ao acaso dos dois pesquisadores terem excluído os mesmos estudos, ou seja, $Ca(E)$. Para isso, multiplicam-se os marginais totais da linha (E) e da coluna (E) (conforme ilustrado na Figura 4.2(c)). Dessa forma, $Ca(E)$ é 20% (que corresponde à 40% * 50%). Em seguida, o valor da concordância ao acaso Ca é calculado somando-se ambas as concordâncias

ao acaso, resultando em 50% (que corresponde a $Ca(I) + Ca(E) = 30\% + 20\%$). O valor da Kappa é então calculado utilizando-se a expressão matemática, o que resulta em Kappa igual a 40% ou 0,4 (conforme ilustrado na Figura 4.2(d)).

Para interpretar o valor do Kappa, pode-se adotar a tabela de concordância proposta por Landis e Koch [84]. Como regra, valores do Kappa menores do que zero são considerados sem concordância, valores de 0 a 0,19 (concordância pobre), de 0,20 – 0,39 (concordância considerável), de 0,40 a 0,59 (concordância moderada), de 0,60 a 0,79 (concordância substancial) e de 0,80 a 1,00 (concordância excelente). Quanto mais próximo de 1, maior a confiabilidade da seleção. Por outro lado, valores próximos ou menores que zero sugerem que a concordância foi ao acaso. No nosso exemplo, a concordância foi moderada. É recomendado que os valores do Kappa sejam maiores do que 0,6. Exemplos de RSs na área de ES que fizeram uso do teste Kappa são os trabalhos de Kitchenham e Brereton [63] e Osborne-O'Hagan et al. [85].

		Pesquisador 1		$Co = 40\% + 30\% = 70\%$
		(I)	(E)	
Pesquisador 2	(I)	40%	20%	
	(E)	10%	30%	

(a) Cálculo da Concordância observada (Co)

		Pesquisador 1		Total marginal da linha (TML)
		(I)	(E)	
Pesquisador 2	(I)	40%	20%	60%
	(E)	10%	30%	40%
Total marginal da coluna (TMC)		50%	50%	

(b) Cálculo dos Totais Marginais da Linha (TML) e Totais Marginais da Coluna (TMC)

$$Ca(I) = TML(I) * TMC(I) = 60\% * 50\% = 30\%$$

$$Ca(E) = TML(E) * TMC(E) = 40\% * 50\% = 20\%$$

$$Ca = Ca(I) + Ca(E) = 30\% + 20\% = 50\%$$

(c) Cálculo da Concordância ao acaso (Ca)

$$Kappa = \frac{Co - Ca}{100\% - Ca} = \frac{70\% - 50\%}{100\% - 50\%} = 40\% = 0,4$$

(d) Cálculo do Kappa

Figura 4.2: Etapas do cálculo do teste Kappa.

Questão 5: Qual é o número adequado de estudos a serem incluídos em uma RS?

Não há um número mais adequado de estudos a serem incluídos em uma RS. Na literatura, é possível encontrar revisões que incluíram em torno de 50% dos estudos identificados, até aquelas que consideraram relevantes menos de 1% do total de estudos analisados. Alguns exemplos a seguir ilustram essa variação.

- [86] dos 343 candidatos identificados, 163 foram incluídos – 47,52%;
- [39] dos 185 candidatos identificados, 38 foram incluídos – 20,84%
- [87] dos 379 candidatos identificados, 74 foram incluídos – 19,52%
- [75] dos 208 candidatos identificados, 36 foram incluídos – 17,30%
- [88] dos 1.034 candidatos identificados, 83 foram incluídos – 8,32%
- [82] dos 564 candidatos identificados, 26 foram incluídos – 4,60%
- [89] dos 3.193 candidatos identificados, 81 foram incluídos – 2,53%
- [90] dos 5.453 candidatos identificados, 103 foram incluídos – 1,92%
- [42] dos 15.430 candidatos identificados, 196 foram incluídos – 1,27%

De fato, o percentual de estudos incluídos em uma RS está bastante relacionado à *string* de busca utilizada. Se a *string* de busca conseguir ser precisa em relação ao tema da RS, então é provável que o número de estudos candidatos identificados seja menor e que o percentual de estudos incluídos seja mais alto. Como mostrado nos exemplos, normalmente, um baixo percentual de inclusão é acompanhado de um elevado número de estudos identificados como candidatos, enquanto um percentual mais elevado de estudos incluídos é associado a um número menor de estudos identificados como candidatos. Contudo, nem sempre é possível calibrar a *string* de busca para obter elevado percentual de inclusão, sob pena de, ao se usar uma *string* de busca mais restritiva, deixar de fora estudos relevantes. Para mais detalhes sobre a composição de *strings* de busca, consulte o Capítulo 3.

Questão 6: A seleção deve ser documentada?

Toda a seleção deve ser documentada e essa documentação pode ser realizada com o auxílio, por exemplo, de planilhas. Na Figura 4.3 é apresentada uma proposta de documentação para a seleção de estudos.

Inicialmente (veja Figura 4.3 – Planilha 1), todos os estudos retornados da atividade de busca, sem eliminar as repetições (estudos presentes em mais de uma fonte), devem ser catalogados, incluindo, pelo menos, as seguintes informações: título (título do estudo), autores (autores do estudo), ano (ano de publicação do estudo) e base (fonte de pesquisa do estudo). Nessa planilha, os estudos devem ser ordenados por base. Para fins de exemplificação, a planilha da

Figura 4.3 foi hipoteticamente preenchida para simular uma atividade de seleção. Nesse exemplo, observe que os estudos das linhas 2 e 3, retornados da mesma base de dados (Scopus), estão ordenados de forma sequencial (do maior para o menor ano).

A seguir, os estudos repetidos devem ser eliminados, deixando registrado para cada estudo que estava repetido todas as bases que o retornou (veja Figura 4.3 – Planilha 2). Para cada estudo, deve ser atribuído um identificador (ID) único. Assim, na Planilha 2 devem constar todos os estudos a serem avaliados durante a condução da seleção inicial. No exemplo, considerando que os estudos das linhas 1, 4 e 5 da Planilha 1 são repetidos, apenas uma de suas ocorrências permanece na planilha (ID1 na Planilha 2), sendo registradas todas as bases em que ele está presente.

PLANILHA 1: Todos os estudos						
	Título	Autores	Ano	Base		
1	Título do estudo 1	Primeiro autor et al.	2015	IEEE xplore		
	Título do estudo 2	Primeiro autor et al.	2014	Scopus		
	Título do estudo 3	Primeiro autor et al.	2010	Scopus		
	Título do estudo 4	Primeiro autor et al.	2015	Web of Science		
	Título do estudo 5	Primeiro autor et al.	2015	ISI of Knowledge		
	Título do estudo 6	Primeiro autor et al.	2013	Science Direct		
	Título do estudo 7	Primeiro autor et al.	2014	ACM Digital Library		

PLANILHA 2: Seleção Inicial							
ID	Título	Autores	Ano	Bases	Resumo	Análise	Revisão da Análise
1	Título do estudo 1	Primeiro autor et al.	2015	IEEE Xplorer Web of Science ISI of Knowledge	Resumo estudo 1	Excluído (CE 1)	OK
2	Título do estudo 2	Primeiro autor et al.	2014	Scopus	Resumo estudo 2	Incluído (CI 1)	Excluído (CE3) – Registro revisão
3	Título do estudo 3	Primeiro autor et al.	2010	Scopus	Resumo estudo 3	Incluído (CI 1)	
4	Título do estudo 6	Primeiro autor et al.	2013	Science Direct	Resumo estudo 6	Incluído (CI 2)	Idem
5	Título do estudo 7	Primeiro autor et al.	2014	ACM Digital Library	Resumo estudo 7	Incluído (CI 1)	

PLANILHA 3: Seleção Final						
ID	Título	Autores	Ano	Bases	Análise	Revisão da Análise
3	Título do estudo 3	Primeiro autor et al.	2010	Scopus	Incluído (CI 1)	OK
4	Título do estudo 6	Primeiro autor et al.	2013	Science Direct	Excluído (CE 2)	OK
5	Título do estudo 7	Primeiro autor et al.	2014	ACM Digital Library	Incluído (CI 1)	OK

PLANILHA 4: Resultado Final					
ID	Título	Autores	Ano	Bases	Referência completa
3	Título do estudo 3	Primeiro autor et al.	2010	Scopus	Referência estudo 3
5	Título do estudo 7	Primeiro autor et al.	2014	ACM Digital Library	Referência estudo 7

Figura 4.3: Sugestão para a documentação da seleção de estudos.

Na sequência, cada estudo deve ser analisado, levando em conta os critérios de seleção (critérios de inclusão e exclusão) aplicados considerando a leitura apenas do título e resumo. Caso a análise aponte para a exclusão do estudo, na coluna *Análise* deve ser indicado o critério de exclusão aplicado. Nesses casos, pode ser interessante não só listar o critério utilizado, mas também as razões para a sua exclusão. De maneira similar, caso a análise aponte para a inclusão do estudo, na coluna *Análise* deve ser indicado o critério de inclusão associado.

Além disso, o resultado da análise pode ser revisto por outros pesquisadores. O resultado dessa revisão deve ser registrado na coluna *Revisão da Análise*. Estudos em que haja

divergência de opinião por parte dos pesquisadores devem ter a coluna *Revisão da Análise* com a classificação do segundo revisor. Estudos que não apresentaram divergência de opinião por parte dos pesquisadores devem ter a coluna *Revisão da Análise* com a marcação *OK* ou *Idem*. Por fim, como nem todos os estudos terão sua análise revisada, aqueles não selecionados para revisão terão a coluna *Revisão da Análise* em branco. No exemplo da Figura 4.3 – Planilha 2, os estudos ID1 e ID4 foram revisados, mas não houve divergência de opinião entre os pesquisadores. Os estudos ID3 e ID5, por sua vez, não tiveram a análise revisada. Por fim, o estudo ID2 teve avaliações divergentes. Estudos em que haja divergência de opinião por parte dos pesquisadores devem ser objeto de discussão, procurando se chegar a um consenso. O resultado final deve ser registrado em uma nova coluna (não mostrada na Figura 4.3).

Após a condução da seleção final na qual os estudos foram lidos na íntegra e selecionados usando os critérios de seleção, o registro da seleção final é realizado na Planilha 3 (veja Figura 4.3), de maneira análoga ao realizado na seleção inicial usando os campos *Análise* e *Revisão da Análise*. No exemplo da Figura 4.3, todos os estudos foram revisados e não houve divergências.

Por fim, cada estudo selecionado durante a seleção final é registrado na planilha 4 (veja Figura 4.3). Nessa planilha são registrados os estudos finais, sendo que desse conjunto de estudos, deverão ser extraídos os dados e analisadas as questões de pesquisa. No exemplo, dois estudos foram incluídos: ID3 e ID5.

Considerações finais

A seleção de estudos primários, que leva em conta critérios de inclusão, exclusão e de qualidade, está diretamente relacionada à qualidade dos resultados de uma RS. Assim, o foco deste capítulo foi apresentar as etapas da seleção, desde a etapa inicial até a etapa final, com a definição dos estudos considerados relevantes para terem seus dados extraídos e sumarizados. Para compreender como sintetizar, resumir e apresentar os resultados, leia o Capítulo 5. Em resumo, os principais pontos de aprendizagem deste capítulo foram:

1. A seleção de estudos é inicialmente realizada por meio da aplicação dos critérios de seleção após a leitura do título e resumo dos estudos. A confirmação da inclusão dos estudos é feita após a leitura do texto completo;
2. Visando aumentar a confiabilidade da seleção, essa deve ser conduzida preferencialmente por dois ou mais pesquisadores, de forma independente, ou por um pesquisador com um outro revisando pelo menos 1/3 dos estudos retornados. As divergências devem ser solucionadas por consenso ou por um terceiro pesquisador;

3. Por meio da avaliação da qualidade, é possível analisar mais detalhadamente aspectos metodológicos dos estudos. Incluir estudos de qualidade é essencial para garantir a qualidade da RS; e
4. A seleção deve ser documentada visando transparência, auditabilidade e, eventualmente, sua reprodução.

5. Síntese dos dados e apresentação dos resultados

Elisa Yumi Nakagawa
Milena Guessi
Lucas Bueno Ruas de Oliveira
Brauner Roberto do Nascimento Oliveira

O objetivo de uma RS é responder às questões de pesquisa com base nos estudos primários que foram selecionados como relevantes para aquela RS, iniciando-se então pela atividade de análise desses estudos. Durante a condução da análise, deve-se ter como principal preocupação o estabelecimento de uma ligação lógica entre os estudos incluídos e as respostas às questões. Para isso, deve-se realizar a extração dos dados gerais e específicos de cada estudo primário, contando preferencialmente com o suporte automatizado de ferramentas de software. Utilizando-se de métodos qualitativos, quantitativos ou mistos para a síntese dos dados extraídos, esses dados são sumarizados para responder às questões de pesquisa. Nesse ponto, avaliar a qualidade e heterogeneidade dos estudos primários sendo considerados é essencial quando se visa obter resultados que retratem mais fielmente o tópico de pesquisa investigado.

Destaca-se ainda o cuidado que deve ser dado a uma adequada documentação dos resultados da RS, visando assegurar a confiabilidade nas conclusões, bem como facilitar sua disseminação. Em particular, diferentes meios para disseminar uma RS podem ser empregados, dependendo do público alvo e do propósito da revisão.

Salienta-se que a área de ES, assim como as diversas áreas da Computação, evolui continuamente, o que faz com que as RSs conduzidas nessa área precisem ser atualizadas de tempos em tempos. Para isso, existem indicadores que podem ser considerados para decidir quando uma RS precisa ser atualizada. Neste capítulo serão então abordados os principais fatores a serem considerados para sintetizar os dados, apresentar os resultados e também atualizar os resultados de uma RS.

Questão 1: O que é a extração de dados?

A extração de dados é uma atividade na qual são obtidos dados a partir da leitura completa dos estudos primários incluídos na etapa de seleção. O objetivo dessa atividade é registrar de maneira precisa os dados necessários para responder às questões de pesquisa estabelecidas para a RS [7]. É também possível que a extração ocorra em paralelo à atividade de seleção [53].

Os primeiros dados a serem coletados de cada estudo primário referem-se aos metadados que serão utilizados para identificar de forma única cada estudo e também gerenciar a atividade de extração. Dados referentes aos seguintes metadados podem ser extraídos [7]: título, autores, veículo de publicação (por exemplo, nome do evento ou periódico), nome do(s) revisor(es) e data da extração. Também pode ser útil considerar outros metadados, tais como o ano de publicação, a afiliação dos autores, entre outros. A decisão sobre quais metadados considerar depende dos objetivos de cada RS.

Após a extração desses primeiros dados, deve-se coletar dados que sejam relevantes para responder às questões de pesquisa. Para cada questão, devem ser coletados dados referentes aos diferentes campos estabelecidos no formulário de extração de dados, definido inicialmente no protocolo da revisão. Em particular, o conjunto de dados a serem extraídos depende diretamente das questões de pesquisa. Por exemplo, a RS conduzida por Kitchenham et al. [4] teve como uma das questões de pesquisa analisar se, após as recomendações iniciais de 2004 [6] [10], novas diretrizes para a condução de estudos secundários haviam sido propostas. Para responder a essa questão, os seguintes dados foram extraídos: (i) o tipo do estudo secundário (por exemplo, RS ou MS); e (ii) se o estudo propunha novas diretrizes para a condução de estudos secundários e também apresentava lições aprendidas. Desse modo, a resposta para a questão em particular foi elaborada a partir da sumarização dos estudos secundários que propuseram melhorias nas diretrizes existentes. Nota-se que há uma ligação lógica entre os dados coletados e a resposta à questão.

Em alguns casos, dados relevantes sobre um estudo primário podem não estar presentes ou aparecem de forma pouco detalhada. Para atenuar esse problema, é possível enriquecer a extração de dados por meio de uma busca por materiais complementares relacionados ao estudo em questão, tais como relatórios técnicos, dissertações, teses e sites web [7] [8]. Normalmente, esses materiais estão referenciados no próprio estudo ou mesmo nas páginas pessoais dos autores. Um outro meio de se obter materiais e informações adicionais sobre o estudo é contactar diretamente os próprios autores [7].

Questão 2: Como distribuir os estudos incluídos entre os revisores para realizar a extração de dados?

É recomendável que a extração de dados seja conduzida por mais de um revisor, visando aumentar a confiabilidade nos dados que serão posteriormente utilizados para responder às questões de pesquisa. Dados extraídos de um mesmo estudo por dois ou mais revisores de forma independente podem ser comparados para que divergências possam ser identificadas e

resolvidas ao final da atividade de extração. Para minimizar a ocorrência de divergências, bem como problemas durante a extração, é fundamental que sejam realizados um ou mais testes piloto da atividade de extração, ainda durante a elaboração do protocolo. Alguns problemas que podem ser encontrados antecipadamente ao realizar o teste piloto são [9]:

- Os envolvidos na extração de dados podem não ter o mesmo entendimento do conteúdo a ser extraído que os responsáveis pela elaboração do formulário de extração;
- Os estudos primários podem reportar seus resultados de diferentes maneiras, o que, por vezes, leva à necessidade de alteração na atividade de extração; e
- Os estudos podem não reportar todos os dados necessários para se preencher o formulário de extração, o que demanda a definição de uma estratégia para lidar com a ausência de dados.

A escolha da configuração da equipe de revisores para a atividade de extração de dados deve levar em consideração o conjunto de revisores disponíveis, a quantidade de estudos primários incluídos durante a seleção e o tempo disponível para a conclusão da RS. Algumas das possíveis configurações são [7]:

- **Um revisor e um avaliador:** Um revisor realiza a extração dos dados de todos os estudos, enquanto que o avaliador conduz a extração de dados de uma amostra aleatória. Em seguida, os dados extraídos por ambos são confrontados de forma a identificar divergências. Considerada a configuração mais simples, envolve geralmente um aluno de pós-graduação no papel de revisor e um supervisor (orientador ou especialista no tópico da pesquisa da RS) como o avaliador. Um exemplo de RS que utilizou essa configuração é encontrado em [87].
- **Dois revisores independentes:** A extração de dados de todos os estudos primários é realizada por dois revisores de forma independente. Em seguida, os dados são comparados e divergências entre os revisores são resolvidas. Caso os revisores não cheguem em um acordo sobre algum dado extraído, um avaliador externo pode ser consultado para resolver a divergência. Um exemplo de RS que utilizou essa configuração é encontrado em [91]. Nos casos em que a RS envolve a coleta de dados de muitos estudos primários, a extração por dois revisores independentes pode se tornar demorada e pouco produtiva.
- **Dois ou mais revisores não-independentes:** O conjunto de estudos primários é dividido entre todos os revisores, sendo que cada estudo terá seus dados extraídos por

apenas um revisor. Essa configuração é interessante quando há muitos estudos e há pouco tempo para a conclusão da atividade de extração. Para ter ideia da confiabilidade dos dados coletados, pelo menos um subconjunto dos estudos deve ter seus dados extraídos por todos os revisores, para verificar se apresentam divergências. Caso os dados extraídos por dois ou mais revisores não apresentarem divergências, pode-se de alguma forma assumir que o restante dos dados coletados são confiáveis. Um exemplo de RS que utilizou essa configuração pode ser visto em [92].

Questão 3: Como gerenciar e armazenar os dados extraídos?

Durante a atividade de extração, em geral, muitos dados são coletados. Para facilitar a coleta, armazenamento, organização e posterior interpretação e síntese desses dados, ferramentas de software comumente utilizadas no gerenciamento de referências bibliográficas, tais como o *Mendeley*¹⁹, *Zotero*²⁰ e *JabRef*²¹, podem ser úteis. A ferramenta *JabRef*, por exemplo, permite que novos campos sejam criados além daqueles inicialmente definidos para arquivos de bibliografia no formato *BibTeX*²², o que possibilita incluir os campos estabelecidos no formulário de extração de dados. As planilhas eletrônicas são uma outra alternativa, fornecendo também suporte para a coleta, armazenamento, organização, classificação e síntese de dados, além de possibilitar a geração de gráficos que facilitam a visualização dos dados. Já os formulários *online* fornecem as funcionalidades das ferramentas de gerenciamento de referências, além do suporte fornecido pelas planilhas eletrônicas. Em particular, formulários *online* possibilitam que os dados extraídos sejam padronizados, uma vez que permitem estabelecer um conjunto de valores admissíveis para cada campo de extração. Além disso, por serem comumente disponibilizadas por meio de um navegador web, permite a distribuição da atividade de extração entre os diversos revisores. Formulários *online*, tais como o *Google Form*²³, podem produzir gráficos para a sumarização de dados, além de permitirem a exportação dos dados no formato de planilhas eletrônicas.

Tanto as ferramentas de gerenciamento de referências quanto as planilhas eletrônicas e formulários *online* são ferramentas genéricas que, apesar de muito úteis, não foram projetadas especificamente para dar suporte à execução de RSs. Nesse cenário, diferentes iniciativas focadas na automatização da condução de RSs podem ser encontradas.

¹⁹ <http://www.mendeley.com>

²⁰ <http://www.zotero.org>

²¹ <http://jabref.sourceforge.net>

²² <http://www.bibtex.org/Format>

²³ <https://www.google.com/forms>

Especificamente para a atividade de extração de dados, existem duas ferramentas principais: *Projection Explorer – Pex*²⁴ e *UNITEX*²⁵. Para apoiar o processo completo da RS, as quatro ferramentas mais conhecidas são: *SLR-Tool*²⁶, *SLuRp*²⁷, *StArt*²⁸ e *SLRTOOL*²⁹. De um modo geral, as quatro ferramentas permitem a criação e o uso de formulários de extração de dados, bem como o apoio à síntese dos dados. Diferenças entre elas existem, por exemplo, enquanto a *SLuRp* permite a criação de dois formulários distintos, um para a extração de dados qualitativos e outro para a extração de dados quantitativos, a *StArt* utiliza somente um formulário. Além disso, vale destacar que todas essas seis ferramentas são de acesso gratuito.

Nossa experiência com a execução de diversas RSs mostra que a utilização de duas ou mais ferramentas em conjunto também é uma solução produtiva na execução de RSs. Além disso, tem-se observado que não existe ferramenta melhor ou pior, mas sim a mais adequada a cada equipe responsável pela RS, à configuração da equipe de revisores, à familiaridade dos revisores com as ferramentas, aos tipos de estudo sendo avaliados e também aos recursos disponíveis. Portanto, não é possível definir qual ou quais ferramentas podem aliar maior produtividade na execução de RSs.

Questão 4: Como sintetizar os dados extraídos?

A atividade de síntese é essencial em uma RS e tem como objetivo principal combinar os dados extraídos de cada um dos estudos primários considerados. Diferentes métodos de síntese podem ser empregados para estabelecer uma estrutura lógica entre os estudos primários e as conclusões geradas pela RS [93]. A escolha por um determinado método deve ser pautada pelas características da RS, tais como o tipo dos estudos primários (por exemplo, experimentos controlados, estudos de caso e entrevistas), o tipo de dados extraídos (por exemplo, textos, figuras e números) e também pela quantidade de estudos disponíveis (por exemplo, dezenas, centenas ou milhares). A seguir, apresentaremos alguns dos métodos que têm sido empregados para sintetizar evidências na área de ES.

Métodos quantitativos são empregados para sintetizar dados provenientes de experimentos controlados, estudos de caso ou estudos observacionais que buscam detectar a existência de um efeito, por exemplo, se o uso de um método de projeto ou ferramenta irá aumentar a

²⁴ <http://infoserver.lcad.icmc.usp.br/infovis2/PEx>

²⁵ <http://www-igm.univ-mly.fr/~unitex>

²⁶ <http://alarcosj.esi.uclm.es/SLRTool>

²⁷ <https://bugcatcher.stca.herts.ac.uk/slurp/>

²⁸ http://lapes.dc.ufscar.br/tools/start_tool

²⁹ <http://www.slrtool.org>

produtividade ou quais fatores impactam a qualidade de um produto [94]. Desse modo, além de confirmar a existência de um efeito, esse tipo de análise também permite estimar a magnitude de um efeito, que é referenciada como *tamanho do efeito*. Para isso, a meta-análise emprega métodos estatísticos que permitem combinar dados numéricos reportados em vários estudos independentes [8] [94] e, assim, verificar se os resultados empíricos são consistentes entre diferentes estudos [94]. Os passos de uma meta-análise são: (i) decidir quais estudos incluir na síntese; (ii) estimar o tamanho do efeito em cada estudo; e (iii) combinar as estimativas individuais para estimar e testar o efeito combinado. A adoção desse método de síntese assume a homogeneidade dos estudos incluídos na análise, ou seja, os estudos são do mesmo tipo, possuem o mesmo teste de hipótese, consideram as mesmas medidas para os efeitos e, além disso, relatam os mesmos fatores explicativos [94]. Uma vez que essa etapa seja concluída com sucesso, pode-se combinar as estimativas individuais para determinar o tamanho médio de um efeito. Caso os critérios de homogeneidade sejam confirmados, pode-se utilizar o modelo de efeito fixo para o tamanho do efeito médio. Caso contrário, o modelo de efeitos randômicos pode ser empregado uma vez que também estima a variabilidade causada por um fator desconhecido [94]. Por exemplo, em [95], estudos empíricos e quantitativos são sintetizados para comprovar se podem ser observadas melhorias no desenvolvimento e manutenção de software nas empresas que investiram em programas de qualidade. A combinação desses estudos permite confirmar o impacto positivo de programas de qualidade mesmo que nenhum dos estudos incluídos tenha sido conclusivo sobre essa questão.

Vale destacar que a abrangência de tópicos tratados em ES, tais como processos, sistemas, atividades, tecnologias e linguagens, faz com que diferentes tipos de estudos e de evidências precisem ser utilizados para responder às questões de pesquisa. Por exemplo, estudos nessa área podem ter como objetivo compreender o comportamento humano no processo de desenvolvimento de software, como motivação, percepção e experiência [96]. Nesse cenário, vários estudos relatam dificuldades em empregar a meta-análise em RSs na área de ES, isso porque os estudos costumam ser qualitativos e heterogêneos, poucos estudos reportam uma estimativa para o tamanho do efeito e demais informações sobre os dados observados (por exemplo, média, variância e normalidade) são omitidas [9] [97]. Desse modo, métodos qualitativos podem ser utilizados para extrair tendências, padrões e generalizações de evidências que foram obtidas em estudos de tipos diferentes [46] [96]. O resultado da síntese qualitativa pode ser rico e explicativo por levar em consideração o contexto em que as evidências foram observadas [96]. Esses métodos podem ser diferenciados entre [98]: (i)

métodos integrativos, que visam sumarizar e agrupar evidências, e (ii) métodos interpretativos, que buscam organizar os conceitos identificados nos estudos incluídos em uma teoria que seja abrangente a todos eles, ou seja, que possa explicar todos os estudos incluídos. Em ambos os casos, os métodos qualitativos buscam estabelecer uma conexão forte entre as evidências (do inglês, *chain of evidence*) e os resultados para viabilizar a replicação e confirmação da análise por revisores externos. Recentemente, tem-se discutido também a adoção de métodos de síntese mista, que visam sintetizar evidências qualitativas e quantitativas [93] [96]. A seguir, sumarizamos os métodos de síntese mista que têm sido empregados nas RSs em ES.

A síntese narrativa constrói uma história a partir das evidências encontradas nos estudos incluídos. Os passos recomendados para conduzir a síntese narrativa são [99]: (i) desenvolvimento de uma teoria; (ii) desenvolvimento de uma síntese preliminar; (iii) exploração de relacionamentos dentro e entre estudos; e (iv) avaliação da robustez do produto da síntese. No primeiro passo, uma possível teoria para os objetivos, contexto e resultados dos estudos é criada. Em seguida, os estudos primários podem ser tabulados, agrupados, codificados e/ou traduzidos³⁰. A partir desse ponto, é possível explorar relacionamentos dentro e entre os estudos incluídos, tais como a conexão entre as características de um estudo e os seus resultados e a conexão entre os resultados de diferentes estudos. Para apoiar essa atividade, pode-se adotar a triangulação³¹ conceitual, do pesquisador e/ou metodológica, construir uma descrição qualitativa que sumariza as contribuições de cada estudo ou representar graficamente as conexões entre características e resultados. Por fim, a validade das conclusões geradas pela síntese narrativa pode ser fortalecida considerando-se, por exemplo, apenas as evidências de melhor qualidade, reflexão crítica sobre o processo de síntese e/ou também pela confirmação dos resultados da síntese com os autores dos estudos que foram considerados. Por exemplo, na RS apresentada por Afzal et al. [100], a síntese narrativa é empregada para descrever cronologicamente as evidências obtidas em cada estudo e discutir as diferenças e possíveis explicações para os resultados.

A síntese comparativa qualitativa analisa um conjunto de relações causais utilizando lógica booleana e, por meio de uma tabela verdade, explica os caminhos que levam até um resultado específico [101]. Assim, utilizam-se as combinações de variáveis independentes e resultados

³⁰ Tradução nesse contexto refere-se à integração de temas e conceitos relacionados que foram reportados por estudos diferentes.

³¹ A triangulação emprega mais de um método ou perspectiva para capturar um determinado aspecto de interesse.

reportados pelos estudos incluídos para identificar as condições necessárias e suficientes de um resultado [102]. É possível minimizar essa tabela verdade para eliminar combinações de variáveis que sejam logicamente inconsistentes e, desse modo, restringir o número de combinações. De qualquer modo, esse método de síntese possibilita a exploração de padrões complexos e variados sobre a causa de um determinado resultado. Embora o conceito de tabela verdade e álgebra booleana não seja explicitamente aplicado, Davis et al. [103] desenvolveram uma RS em que uma tabela foi construída para comparar estudos que forneciam evidências a favor ou contra um determinado resultado obtido pela agregação de dados dos estudos incluídos, além de relatar quaisquer dificuldades encontradas durante essa agregação. Já Kitchenham et al. [7] apresentaram tabelas comparando a aplicação de modelos de estimação em dois contextos distintos. Em seguida, eles consideraram vantagens e limitações de cada abordagem e verificaram as evidências trazidas por cada estudo. Por fim, a partir do resultado, foi criada uma tabela das recomendações a favor ou contra cada um.

Embora seja bastante comum em estudos que foram publicados como RS [93], a análise de escopo (do inglês, *scoping analysis*) é um método descritivo. Em vez de sintetizar os resultados dos estudos incluídos, essas RSs apresentam apenas um panorama das pesquisas existentes sobre um determinado tópico. Por essa razão, um estudo que relata apenas esse tipo de análise deve ser referenciado como um MS [8]. De fato, a falta de síntese e de avaliação da qualidade dos estudos incluídos é o que diferencia um MS de uma RS [93].

Considerando-se os métodos de síntese discutidos nessa questão, a escolha por um ou outro deve ser pautada pelas características da RS, tais como, o tipo de estudos incluídos (ou seja, experimentos, estudos de caso, etc.), o tipo de evidência e o tipo de questão de pesquisa. Em particular, a quantidade de estudos disponíveis (por exemplo, dezenas ou centenas) também pode determinar a viabilidade de se conduzir a síntese utilizando um método específico. De fato, cada método de síntese é particular com relação aos pontos fortes e fracos, capacidade para lidar com evidências quantitativas e qualitativas e tipos de questão para os quais eles são mais adequados [102]. É importante ressaltar que não há consenso para algumas limitações apresentadas por métodos qualitativos e mistos, como a transformação de dados qualitativos em quantitativos [102]. A Tabela 5.1 sumariza as características dos métodos de síntese apresentados anteriormente e exemplifica as questões de pesquisa que já foram respondidas com eles.

Tabela 5.1: Comparação entre métodos de síntese.

Métodos de Síntese				
	Meta-análise	Narrativa	Temática	Comparativa Qualitativa
<i>Pontos Fortes</i>	Permite afirmar com uma margem de segurança a existência ou não de um efeito e seu tamanho médio [94]	Pode construir explicações ricas que incluem comentários e abstrações de alto nível [102]	Segue um processo de síntese flexível e permite sintetizar diferentes tipos de estudos [102]	Segue um processo de síntese transparente, sistemático e lógico; encoraja uma abordagem evolucionária e integrativa do conhecimento ao permitir a síntese de estudos secundários e primários [102]
<i>Pontos Fracos</i>	Falta de informação sobre os estudos e os dados dos experimentos podem dificultar a adoção do método; homogeneidade e qualidade dos estudos incluídos impactam a confiabilidade da síntese [94]	Segue um processo de síntese informal que pode ser criticado pela falta de transparência [102]	Se considerar apenas os temas reportados pelos estudos incluídos, restringirá os temas de ordem mais elevada que podem aparecer [102]	Depende da transformação de evidências qualitativas em quantitativas; considera a ausência de evidência como evidência da ausência [102]
<i>Tipo de evidências</i>	Quantitativa	Quantitativa e Qualitativa	Quantitativa e Qualitativa	Quantitativa e Qualitativa
<i>Exemplo de pergunta</i>	O uso de uma nova ferramenta melhora a produtividade de desenvolvimento (sem nenhum efeito prejudicial na qualidade) em comparação com o uso de uma ferramenta existente? [94]	Em quais áreas de teste não-funcional têm sido aplicadas meta-heurísticas de busca? [100]	Quais são os desafios de coordenação apresentados por dimensões de dispersão nos resultados de projetos globais de desenvolvimento de software? [104]	A RS tem como objetivo comparar os resultados de estudos empíricos sobre a eficácia de diferentes técnicas para elicitação de requisitos [103]

Nosso objetivo aqui foi apresentar os principais métodos de síntese que têm sido empregados nas RSs da área de ES. Outros tipos de síntese também podem ser empregados, tais como contagem de votos (do inglês, *vote-counting*) [94], teoria fundamentada em dados (do inglês, *grounded theory*) [105] [106], meta-ethnografia (do inglês, *meta-ethnography*) [98], síntese cruzada (do inglês, *cross-case*) [107] e síntese realista (do inglês, *realist synthesis*) [108]. O leitor poderá consultar essas referências para recomendações específicas sobre cada método, além de discussões detalhadas sobre suas diferenças, vantagens e limitações em [93] [96] [102] [109], o que poderá fundamentar a decisão pela aplicação de um ou outro método em uma dada RS.

Por fim, vale salientar que na documentação de uma RS é importante deixar registrado o método de síntese empregado e as recomendações que foram seguidas, uma vez que permitirá a avaliação e replicação do processo de síntese e, assim, aumentar a confiabilidade dos resultados obtidos [93].

Questão 5: O que é a análise de sensibilidade, quando e como aplicá-la?

A análise de sensibilidade é um procedimento utilizado para determinar qual a sensibilidade dos resultados de uma RS quando alteradas as premissas iniciais definidas no protocolo, tais como os critérios de seleção de estudos primários [110]. Durante a análise de sensibilidade, os resultados de uma RS são também investigados para verificar se estão sendo indevidamente influenciados por subconjuntos de estudos primários incluídos e que apresentam características particulares [7], possibilitando, dessa forma, avaliar o grau de confiança nos resultados ou suposições feitas para aquela RS. Além disso, se for constatada pouca variação entre os resultados dos diferentes subconjuntos de estudos primários, isso implica que os resultados daquela RS são mais confiáveis. Por outro lado, a identificação de variação entre os resultados dos subconjuntos pode indicar a necessidade de uma análise mais detalhada, visando compreender as possíveis razões para essa variação [8]. Diferentes subconjuntos dos estudos primários podem ser considerados, tais como [7] [8]:

- Estudos primários de tipos específicos, tais como experimentos ou *surveys*;
- Estudos primários de qualidade mais alta;
- Estudos primários que utilizem métodos experimentais específicos;
- Estudos primários que durante a seleção ou extração de dados apresentaram divergências entre os envolvidos na RS;
- Estudos primários desenvolvidos em um contexto específico, por exemplo, na indústria; e

- Estudos primários publicados em veículos indexados.

A análise de sensibilidade é aplicável tanto em métodos de síntese qualitativa quanto quantitativa, embora seja mais fácil empregá-la como parte da meta-análise [7]. Quando a síntese da RS é realizada por meio de meta-análise ou os dados quantitativos estão disponíveis, é possível identificar de forma direta estudos primários de qualidade mais alta ou que pertencem a uma classe distinta na hierarquia de tipos (por exemplo, níveis de validação e evidência) [7]. Por outro lado, quando a síntese da RS é qualitativa, a análise de sensibilidade torna-se mais subjetiva. De qualquer forma, é importante verificar, pelo menos, se há impacto significativo na exclusão de estudos de baixa qualidade durante a elaboração das conclusões acerca das questões de pesquisa.

Questão 6: O que é e como relatar os resultados de uma RS?

O relato dos resultados corresponde à fase final de uma RS e envolve as atividades de: (i) documentação dos resultados da revisão; (ii) avaliação da documentação; e (iii) disseminação da revisão para potenciais interessados. Considerando a relevância de conduzir tais atividades de forma efetiva, aconselha-se que a estratégia para documentação e disseminação seja discutida ainda nas fases iniciais da revisão, até mesmo com a inclusão dessa estratégia no protocolo [7]. Além disso, é preciso identificar os possíveis interessados na RS para que o relato seja construído de acordo com as necessidades desses interessados. De um modo geral, os principais interessados em uma RS são os pesquisadores da academia e profissionais da indústria.

É importante salientar que a qualidade de uma RS está diretamente relacionada à qualidade da documentação dos resultados. Portanto, essa documentação deve conter uma reflexão abrangente e, ao mesmo tempo, detalhada dos resultados. Por exemplo, um relato claro sobre o processo de síntese aumenta a confiabilidade da RS, que poderá ser validada e replicada pela comunidade. Bons exemplos de relatos dos resultados de RS podem ser encontrados em [104] [111]. Para apoiar a criação dessa documentação, pode-se utilizar o protocolo da RS para direcionar a discussão dos resultados obtidos de acordo com os objetivos e as questões de pesquisa que precisam ser respondidas [9].

Finalizada a documentação dos resultados, aconselha-se que avaliadores externos à revisão e, se possível, especialistas no tópico de pesquisa, conduzam uma avaliação visando identificar possíveis inconsistências, ambiguidades, limitações, ameaças, superficialidade nas discussões,

entre outros. Essa avaliação costuma ser feita por revisores de periódicos quando essa revisão é submetida a uma revista e por membros do comitê de programa quando submetida a um evento. De qualquer modo, o ideal é que a avaliação ocorra antes mesmo da revisão ser submetida a esses veículos de publicação para aumentar as chances de aceitação. Por outro lado, quando a RS é publicada nos meios que não exigem avaliação para publicação, tais como sites web ou como relatório técnico, a avaliação externa torna-se imprescindível. Vale ainda lembrar que essa avaliação comprehende toda a documentação referente à revisão, o que inclui o protocolo e as informações sobre a condução, a extração de dados, a síntese, além da discussão sobre os resultados da revisão. Para isso, um *checklist* mais geral pode ser encontrado em [7].

Existem diferentes meios para disseminar os resultados de uma RS considerando os diferentes interessados. Para pesquisadores da academia, na grande maioria dos casos, a disseminação tem se dado por meio de periódicos e conferências da área ou subáreas correlatas da ES. Em particular, como periódicos de impacto que têm publicado RSs na área de ES, pode-se citar o *Journal of Systems and Software*³², *Journal of Information and Software Technology*³³, *IEEE Transactions on Software Engineering*³⁴, *Journal Software: Practice and Experience*³⁵ e *Empirical Software Engineering*³⁶. Há também periódicos de subáreas da ES que têm publicado revisões sobre seus temas relacionados, como o *Journal of Requirement Engineering*³⁷, sobre engenharia de requisitos, e *Journal of Software Testing, Verification and Reliability*³⁸, sobre teste de software.

Se a intenção da RS é influenciar profissionais da indústria, outras formas de disseminação podem ser também utilizadas, tais como sites web, revistas e periódicos mais destinados aos profissionais, além de relatórios técnicos disponibilizados em repositórios de documentos.

Embora tenham um bom alcance para a disseminação de RSs, conferências e periódicos trazem, muitas vezes, limitação quanto ao número máximo de páginas. Uma estratégia que pode ser adotada para mitigar tal limitação é complementar o conteúdo dos artigos por meio da disponibilização de informações adicionais em sites web, relatórios técnicos ou mesmo teses de alunos de pós-graduação. Recomenda-se que essa informação adicional conte com o

³² <http://www.journals.elsevier.com/journal-of-systems-and-software>

³³ <http://www.journals.elsevier.com/information-and-software-technology>

³⁴ <http://www.computer.org/web/tse>

³⁵ [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1097-024X](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1097-024X)

³⁶ <http://link.springer.com/journal/10664>

³⁷ <http://link.springer.com/journal/766>

³⁸ [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1099-1689](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1689)

protocolo completo da RS, juntamente com informações sobre as estratégias adotadas para a busca e seleção dos estudos, extração de dados e a síntese do resultados. É também importante apresentar de forma detalhada os dados encontrados, por exemplo, a quantidade e quais foram os estudos primários incluídos e excluídos. Decisões que foram tomadas ao longo da condução da RS devem ser também registradas. Essas informações e dados são cruciais para uma futura atualização da revisão, bem como para a reproduzibilidade da pesquisa em si, caso se faça necessário. Exemplos de RS que usam essa estratégia são [112] [113].

Questão 7: Como estruturar a documentação dos resultados de uma RS?

Organizar adequadamente os resultados de uma RS é fundamental para o seu sucesso. A documentação deverá apresentar de forma clara, objetiva e ao mesmo tempo detalhada o estado da arte sobre o tópico de pesquisa em questão aos possíveis interessados e, com isso, contribuir também para estabelecer um agenda de trabalho, identificando quais linhas de pesquisa ainda estão em aberto. Uma boa estruturação facilita o entendimento de como a revisão foi de fato conduzida, o que promove a confiabilidade nos resultados encontrados e, consequentemente, a disseminação da revisão, trazendo impactos para a área de pesquisa na qual faz parte essa revisão.

Diferentes formatos dessa documentação podem ser encontrados. Observa-se que cada autor tem adotado o que parece ser mais adequado para a sua revisão. Contudo, a grande maioria das revisões reportadas na literatura estruturam a documentação de acordo com as questões de pesquisa estabelecidas no protocolo. Quando observadas as diversas RSs publicadas na área de ES, as discussões sobre os resultados têm abordados os seguintes itens:

- Achados: Apontar os achados mais evidentes, por exemplo, aqueles constatados em diversos estudos primários incluídos ou provenientes de poucos ou mesmo de um estudo primário considerado de alta qualidade (ou seja, que apresentem, por exemplo, resultados com validade estatística). Desse modo, recomenda-se que a qualidade dos estudos primários também seja considerada para diferenciar as contribuições de cada estudo e direcionar a discussão dos resultados da RS. Deve-se também discutir sobre as variações, a magnitude e a aplicabilidade dos achados observados nos estudos primários incluídos;
- Generalizações: Quando permitido pelo método de síntese empregado, é possível generalizar as conclusões sobre um dado achado com base nas evidências encontradas nos estudos primários incluídos;

- Inferências: As discussões podem apresentar inferências baseadas nos achados/dados encontrados nos estudos primários, juntamente com as justificativas de como foram estabelecidas;
- Pontos fracos e fortes: Deve-se abordar os pontos fortes e fracos das evidências encontradas nos estudos primários, o que incluem os benefícios, os efeitos adversos e os riscos, e relacionando-as com resultados encontrados em outras revisões em tópicos vizinhos; e
- Aplicabilidade: Quando permitido pelo método de síntese empregado, é interessante discutir para quem, quando e porque um determinado achado será relevante. Essa informação será especialmente importante para profissionais que irão utilizar os resultados da RS para adotar ou não um novo paradigma, técnica, processo, entre outros.

Tabelas, gráficos e figuras têm sido largamente utilizadas para dar suporte às discussões desenvolvidas nas RSs. Além das discussões minuciosas buscando responder cada questão de pesquisa, vale a pena incluir informações de cunho mais geral que, muitas vezes, têm sido também reportadas em MSs. Esse tipo de informação, que pode ser apresentada inclusive antes das discussões sobre cada questão de pesquisa, pode contribuir para um melhor entendimento do contexto das respostas às questões. Como exemplo, pode-se incluir: (i) distribuição dos estudos primários incluídos ao longo dos anos, o que mostra a tendência do interesse pelo tópico de pesquisa em questão; (ii) distribuição dos estudos incluídos nos diferentes subtópicos abordados; (iii) identificação da origem dos estudos primários, considerando, por exemplo, instituição dos autores (e também se são da academia ou da indústria), grupos de pesquisa ou países de origem; (iv) *word cloud* construído a partir, por exemplo, dos títulos e/ou resumos dos estudos incluídos; e (v) veículos de publicação e qualidade desses veículos segundo o Qualis³⁹ e/ou CORE (*Computing Research and Education Association of Australasia*)⁴⁰, que também pode indicar a relevância desses veículos para o tópico de pesquisa sendo investigado.

Finalizadas as discussões sobre os resultados, deve-se apresentar as limitações e as ameaças à validade da RS como um todo. Baseado nas diversas RSs que temos executado, as principais limitações e ameaças com que temos deparado são: (i) possíveis tendências durante a busca dos estudos primários; (ii) possíveis tendências durante a exclusão de trabalhos primários; (iii) subjetividade durante a extração dos dados a partir dos trabalhos incluídos; e (iv) limitações

³⁹ <http://qualis.capes.gov.br>

⁴⁰ <http://portal.core.edu.au/conf-ranks/>

nas bases bibliográficas que podem não ter retornado todos os trabalhos diretamente relacionados ao tópico de pesquisa da RS. Além dessas, diversas outras ameaças e limitações existem e podem estar presentes em praticamente todas as fases da RS. Para cada ameaça identificada, deve-se apresentar as medidas que foram tomadas para minimizar o seu impacto na revisão. Portanto, explicitar claramente as limitações e ameaças é fundamental, uma vez que permite aos interessados compreender melhor sobre como a revisão foi conduzida, facilitando, assim, uma futura atualização da revisão.

Questão 8: Uma RS precisa ser atualizada?

Após reportar e disseminar uma RS, é importante estar ciente que seu conteúdo precisará ser atualizado futuramente, pois revisões que não são atualizadas tendem a se tornar obsoletas e levar leitores a conclusões equivocadas [110]. A manutenção de uma RS pode ocorrer por dois motivos principais [110]: (i) melhoria do relatório da revisão; e (ii) atualização de conteúdo.

Uma RS precisa de melhorias quando existe a necessidade de alterar o relatório da revisão para, por exemplo, corrigir inconsistências ou reduzir possíveis ameaças à validade dos resultados. Tais melhorias não requerem a recondução da RS, mas podem demandar desde modificações pontuais a esforços consideráveis.

Atualizações de conteúdo envolvem, necessariamente, a busca por novas evidências que possam ser agregadas às existentes em uma RS que já foi concluída [114]. A atualização de conteúdo de uma RS pode envolver outras modificações, como a inclusão ou alteração de questões de pesquisa e critérios de seleção que sejam considerados mais adequados à nova versão da RS. Caso modificações sejam realizadas no protocolo, a busca por estudos deverá ser executada não somente para cobrir o período posterior à última atualização da RS, mas também o período contemplado desde a primeira condução [110]. Se o protocolo não for alterado, a busca por estudos pode resultar na inclusão de novas evidências ou não. Mesmo que não sejam encontrados novos estudos primários, é importante alterar a documentação da RS para registrar informações como a data de realização da última busca e a lista de estudos avaliados [110]. Atualizações de conteúdo que resultem na inclusão de novas evidências devem realizar uma nova síntese dos dados, sumarizar os dados e, se necessário, rever as conclusões relacionadas às respostas de cada uma das questões de pesquisa.

Embora não existam evidências científicas substanciais que permitam determinar de forma eficiente e razoável a necessidade de atualizações em RSs, na área de Medicina há um entendimento de que esse tipo de estudo secundário deve ser revisitado a cada dois anos [110]. Já na área de ES, não há um consenso sobre quando realizar uma atualização em uma

RS. Contudo, mudanças substanciais na área de pesquisa e avanços tecnológicos relevantes podem ser considerados como indicadores de que uma RS precisa ser atualizada. É importante então que os envolvidos na revisão observem as características da área de pesquisa investigada e indiquem uma estimativa de prazo para tais atualizações. RSs que, por algum motivo, não necessitem serem atualizadas devem indicar os motivos para tal decisão. Como exemplos de RSs que sofreram atualizações, podem-se citar a de Oliveira et al. [45], inicialmente conduzida em 2012 e atualizada em 2015 [115], e a de Ferrari e Maldonado [67], atualizada entre os anos de 2006 e 2008 a cada 13 meses.

Ainda que atualizações de RSs seja vitais para manter uma visão realista e atualizada sobre um tópico de pesquisa, há ainda poucos trabalhos na ES destinados a apoiar essa atividade. Ferrari e Maldonado [67] propõem uma alteração no processo descrito por Biolchini et al. [6] para nortear a realização de múltiplas iterações de atualização de uma RS. Dieste et al. [116] complementam o conjunto de atividades para a execução de RSs propostas inicialmente por Kitchenham [10], dando atenção especial às atividades relacionadas à extração, summarização e agregação de evidências encontradas nos estudos primários identificados durante a atualização. Felizardo et al. [117] propõem uma abordagem baseada em análise visual de texto para auxiliar na seleção de novos estudos primários durante a atualização de RSs. A utilização das diretrizes propostas nesses três trabalhos pode servir como um ponto de partida para quem pretende revisitar suas RSs.

Considerações finais

Este capítulo teve como objetivo prover informações que darão subsídios para a realização da síntese e da apresentação dos resultados de uma RS, desde a atividade de extração dos dados, passando pelos diferentes métodos de síntese e formas de documentar e disseminar os resultados e finalizando com uma discussão sobre a necessidade de se manter RS atualizadas. Em resumo, os principais pontos de aprendizagem deste capítulo foram:

- Diferentes métodos têm sido empregados para sintetizar evidências na área de ES, como os métodos quantitativos, qualitativos e os mistos. Sugerimos a leitura de [94] para recomendações sobre meta-análise no contexto de ES. Uma introdução geral sobre métodos qualitativos pode ser encontrada em [118]. Já para uma discussão e referências sobre métodos de síntese qualitativa e mista em ES destacamos a leitura de [93]; e
- RSs precisam ser constantemente atualizadas. Recomendamos a leitura de [67] para uma discussão sobre como atualizar RSs.

6. Mapeamento Sistemático

Ricardo de Almeida Falbo

Érica Ferreira de Souza

Katia Romero Felizardo

RS e MS são tipos de estudos secundários que seguem um processo de pesquisa metodologicamente bem definido para identificar, analisar e interpretar as evidências disponíveis relacionadas a um tópico de pesquisa ou fenômeno de interesse, de uma maneira não tendenciosa e até mesmo repetível [7]. O objetivo deste capítulo é apresentar os principais conceitos sobre MS, destacando suas diferenças com relação à RS. Para ilustrar alguns dos aspectos discutidos neste capítulo, será usado o MS sobre iniciativas de gestão do conhecimento em teste de software publicado em [119].

Questão 1: O que é um MS?

MS é um estudo secundário que tem como objetivo identificar e classificar o conteúdo relacionado a um tópico de pesquisa [7]. O MS corresponde a uma investigação ampla envolvendo estudos primários relacionados ao tópico de pesquisa específico, que visa identificar as evidências disponíveis sobre esse tópico.

Resultados de um MS ajudam a identificar lacunas nessa área, capazes de sugerir pesquisas futuras e prover um guia para posicionar adequadamente novas atividades de pesquisa [7] [120] [121]. Assim, MSs visam prover uma visão geral de um tópico e identificar se há subtópicos nos quais mais estudos primários são necessários.

Questão 2: Por que realizar um MS?

MSs são úteis e trazem vários benefícios em diferentes situações. Para estudantes, tanto de graduação quanto de pós-graduação, MSs ensinam como pesquisar a literatura e organizar os resultados de tais pesquisas. Em especial, estudantes de pós-graduação têm em MSs um meio valioso e eficaz de iniciar suas atividades de pesquisa [15] [122].

Para pesquisadores em geral, incluindo também estudantes, um MS pode prover um corpo de conhecimento e um ponto de partida para suas pesquisas.

Algumas razões típicas para se realizar MSs incluem:

- Para examinar a extensão da pesquisa sobre um dado tópico de pesquisa;
- Para coletar e sumarizar a pesquisa existente sobre um dado tópico de pesquisa. Isso é fundamental para estudantes de pós-graduação, em especial, estudantes de

doutorado iniciando seu trabalho, uma vez que eles têm de compreender o estado da arte da pesquisa no tópico em que eles pretendem trabalhar; e

- Para identificar lacunas existentes em um tópico de pesquisa, que apontem subtópicos promissores para um projeto de pesquisa.

Vale a pena destacar que MSs e RSs são abordagens complementares. Um MS pode preceder uma RS visando prover uma visão geral de um tópico de pesquisa. Ele pode identificar grupos (*clusters*) de estudos que são adequados para estudos mais detalhados e aprofundados, os quais podem ser feitos por meio de RSs [121]. Em uma abordagem na qual um MS precede uma RS, é possível apontar vários benefícios, dentre eles [121]:

- Reduz o tempo necessário para atividades de pesquisa subsequentes, em especial na realização de uma RS;
- Fica mais fácil compreender a literatura e definir questões de pesquisa;
- Provê um conjunto de estudos que pode ser usado como grupo de controle para avaliar a *string* de busca de uma RS; e
- Possibilita que procedimentos, formulários e experiência possam ser reutilizados.

Além disso, um MS pode ser usado [121]:

- Como uma *baseline* para rastrear tendências de pesquisa ao longo do tempo;
- Para justificar a realização de mais estudos primários quando houver poucos ou nenhum estudo empírico relevante;
- Como um meio de identificar estudos relacionados ao trabalho que está sendo desenvolvido pelo pesquisador possibilitando comparação de resultados; e
- Como um recurso educacional para se aprender sobre um dado tópico de pesquisa.

Vale realçar que os benefícios e usos apontados acima só serão obtidos se todas as referências aos estudos selecionados forem citadas e as classificações para cada estudo forem reportadas. Em adição, um MS deve apresentar alta qualidade. É importante que o MS seja baseado em uma busca e seleção rigorosa de estudos, incluindo, por exemplo, busca automática, aplicação de *snowballing* nos estudos selecionados e busca ou comunicação direta com importantes pesquisadores e grupos de pesquisa, bem como em uma classificação feita de maneira confiável e bem definida.

Questão 3: Quais as semelhanças e diferenças entre MS e RS?

De maneira geral, o objetivo de um estudo secundário é prover aos pesquisadores uma visão geral de uma área de pesquisa e ajudar a identificar lacunas na pesquisa nessa área [35]. RSs e

MSs, como estudos secundários, possibilitam a identificação e agregação das evidências disponíveis para responder às questões de pesquisa e auxiliam na identificação de lacunas de pesquisa [15]. Nesse sentido, RSs e MSs apoiam a tomada de decisões relacionadas a uma pesquisa/trabalho a ser desenvolvido e, de fato, há muitas semelhanças entre eles. Contudo, há também diferenças, em especial no que se referem aos objetivos, questões de pesquisa, estratégia de busca, forma de seleção e avaliação da qualidade dos estudos primários, bem como a análise de dados e resultados obtidos [49].

No que se refere aos objetivos, RSs agregam os estudos primários em termos de seus resultados e investigam se esses resultados são consistentes ou contraditórios. RSs visam, portanto, sintetizar evidências [49] [121]. MSs, por sua vez, proveem uma visão mais ampla de um tópico de pesquisa e identificam tanto *clusters* de estudos (que podem ser apropriados para estudos mais detalhados a serem feitos por meio de uma RS), quanto subtópicos em que mais estudos primários são necessários para serem desenvolvidos [123]. MSs estão focados na estruturação de uma área de pesquisa [49] e objetivam classificar os estudos primários relevantes em categorias bem definidas [121].

Segundo Kitchenham et al. [15], RSs agregam resultados relacionados às questões de pesquisa específicas, enquanto MSs visam encontrar e classificar os estudos primários em subtópicos. Os mesmos métodos de busca e extração de dados são usados, mas MSs apoiam-se na tabulação dos estudos primários em categorias. A análise dos resultados foca frequências de estudos por categoria [120].

De fato, as principais diferenças entre MS e RS residem no escopo e nos procedimentos de análise. O escopo de um MS é mais amplo e a análise e síntese mais gerais do que em uma RS [35]. Para tal, a estratégia de busca deve ser menos restritiva, de modo a permitir recuperar mais estudos [121]. Assim, de maneira geral, MSs envolvem mais estudos a serem analisados, enquanto RSs envolvem menos estudos, mas que devem ser analisados com maior profundidade. A Tabela 6.1 sumariza as principais diferenças entre MSs e RSs.

Tabela 6.1: Diferenças entre entre MSs e RSs.

Elementos da RS	Mapeamento Sistemático	Revisão Sistemática
Objetivos	Prover uma visão ampla de um tópico de pesquisa	Prover uma discussão detalhada a partir da sumarização de evidências dos estudos primários
Questões de pesquisa	Genéricas	Específicas
Processo de busca	Definido pelo tópico de pesquisa	Definido pelas questões de pesquisa

Escopo	Amplo	Focado
Estratégia de busca	<i>String</i> de busca mais genérica	<i>String</i> de busca mais específica
Avaliação da qualidade	Não é obrigatória	Recomendável
Resultados	Categorização dos estudos primários de acordo com esquemas de classificação	Discussão detalhada sobre as evidências referentes ao tópico de pesquisa

Questão 4: Quais são as fases do processo de MS?

Por ambos RS e MS serem estudos secundários, o processo discutido no Capítulo 1 (Questão 3), aplica-se tanto a RSs quanto a MSs. Ou seja, o processo de MS envolve as fases de planejamento, condução e publicação dos resultados, executadas de forma iterativa. Contudo, há diretrizes específicas para a realização dessas fases, quando realizadas no contexto de um MS. As diretrizes específicas para MSs serão discutidas neste capítulo.

Vale destacar que todas as informações produzidas nas fases de planejamento e condução devem ser adequadamente registradas para permitir a posterior publicação dos resultados do mapeamento. Para tal, é imprescindível o uso de ferramentas de software para o gerenciamento de referências e registro das informações.

Questão 5: Como definir as questões de pesquisa para um MS?

De maneira geral, as diretrizes discutidas no Capítulo 2 para a descrição das questões de pesquisa aplicam-se tanto às RSs quanto aos MSs. Contudo, uma vez que o propósito de um MS é revisar um tópico mais amplo de pesquisa e classificar os estudos primários relacionados a esse tópico, as questões de pesquisa para um MS são questões mais gerais e, portanto, com menor profundidade. Alguns aspectos tipicamente discutidos em um MS levam a questões, tais como: *Quais subtópicos têm sido tratados? Quais métodos de pesquisa têm sido utilizados? Em que veículos de publicação esses estudos têm sido reportados?* Além disso, MSs têm tipicamente, um número maior de questões de pesquisa se comparados às RSs. Em suma, as questões de pesquisa em um MS são mais gerais e em maior quantidade, na medida que elas buscam descobrir tendências de pesquisa (por exemplo, tendência de publicação ao longo do tempo, tópicos cobertos na literatura, etc).

Petersen et al. [49] apontam cinco aspectos presentes nas questões de pesquisa de muitos dos MSs publicados na literatura, a saber:

- Tipo da contribuição: refere-se a determinar o tipo de intervenção sendo estudado (processo, método, modelo, ferramenta ou métrica);

- Tipo da pesquisa: refere-se ao tipo de estudo apresentado, sendo que uma classificação bastante adotada é a proposta por Wieringa et al. [80], que considera os seguintes tipos de pesquisa: pesquisa de avaliação, pesquisa de validação, proposta de solução, artigo filosófico, relato de experiência e artigo de opinião;
- Foco do estudo: refere-se ao contexto sendo estudado (por exemplo, academia, indústria, governo, projeto ou organização);
- Veículo de publicação: refere-se ao tipo do veículo de publicação nos quais os estudos têm sido publicados (por exemplo, periódicos, conferências e *workshops*); e
- Método de pesquisa: refere-se aos métodos científicos usados (por exemplo, estudo de caso, experimento, *survey*, etc).

Além disso, uma questão de pesquisa sempre presente em MSs é: *Quando os estudos foram publicados?* Essa questão é muito importante, pois mostra o interesse no tópico do MS ao longo do tempo e provê informação que pode ser combinada com outras para extrair tendências ao longo do tempo. Assim, por exemplo, combinando os resultados dessa questão com os resultados de uma questão sobre o tipo de pesquisa, é possível avaliar se a maturidade da pesquisa naquele tópico tem aumentado ao longo dos anos.

Como um exemplo, considere o caso do MS sobre iniciativas de gestão do conhecimento em teste de software, publicado em [119]. O objetivo deste MS é descrever uma visão geral do estado corrente da pesquisa relacionada à gestão de conhecimento aplicada ao teste de software. As questões de pesquisa definidas para esse MS foram as seguintes:

- *QP1: Quando e onde os estudos têm sido publicados?*
- *QP2: Do ponto de vista de teste de software, quais aspectos têm sido enfocados?*
- *QP3: Do ponto de vista de gestão do conhecimento, quais tópicos têm sido focados?*
- *QP4: Quais tipos de pesquisa têm sido realizadas?*
- *QP5: Quais problemas têm sido apontados pelas organizações de software relacionados ao conhecimento sobre teste de software?*
- *QP6: Quais os propósitos de se empregar gestão do conhecimento em teste de software?*
- *QP7: Quais são os tipos de itens de conhecimento tipicamente gerenciados no contexto de teste de software?*
- *QP8: Quais são as tecnologias usadas para prover gestão do conhecimento em teste de software?*

- *QP9: Quais as principais conclusões reportadas relativas à aplicação de gestão do conhecimento em teste de software?*

Questão 6: Qual a estratégia de busca mais adequada para um MS?

A estratégia de busca a ser adotada em um MS é um ponto-chave para o sucesso ou fracasso desse mapeamento. Conforme discutido no Capítulo 3, há três estratégias principais que podem ser aplicadas: busca automática em bases bibliográficas, busca manual e *snowballing*. De maneira geral, para que um mapeamento forneça uma visão ampla de um dado tópico de pesquisa, é necessário realizar a busca automática. A busca manual e *snowballing* são considerados estratégias complementares que devem ser usadas em conjunto com a busca automática. A busca manual deve ser usada quando se sabe que algum evento científico ou periódico específico é um importante veículo de publicação para o tópico em questão e o mesmo não está contemplado nas bases selecionadas para a busca automática. Já *snowballing* é uma estratégia muito útil para encontrar estudos relevantes a partir dos estudos já selecionados por meio da aplicação de outras estratégias ou por meio inclusive de *snowballing*.

Uma quarta estratégia que pode ser aplicada em MSs consiste em, uma vez selecionados os estudos aplicando-se as estratégias anteriormente mencionadas, identificar autores e grupos de pesquisa que têm atuado na área e realizar uma busca por outros estudos desses autores/grupos em suas páginas institucionais. Para tornar essa busca mais sistemática, recomenda-se utilizar o DBLP⁴¹ (*Digital Bibliography & Library Project*), um repositório contendo as principais publicações dos pesquisadores em Ciência da Computação, ou redes sociais de pesquisa como *Research Gate* ou *Academia*.

Vale destacar que aplicar múltiplas estratégias pode ser útil, mas pode consumir muito tempo. Assim, nem sempre é prático aplicar várias estratégias e recomenda-se selecionar as estratégias de forma combinada com as fontes de pesquisa. Se muitas fontes de pesquisa forem usadas, dependendo da quantidade de estudos selecionados, técnicas de busca manual, *snowballing* e pesquisa direta a autores/grupos de pesquisa podem não ser viáveis ou sequer necessárias.

Por exemplo, em [119], inicialmente foi aplicada a busca automática. Contudo, como apenas poucos estudos foram selecionados (13 dos 440 estudos retornados, já eliminados os estudos duplicados), decidiu-se aplicar outras duas estratégias: *snowballing* e pesquisa direta a

⁴¹ <http://dblp.uni-trier.de>

autores/grupos de pesquisa. Aplicando-se *snowballing*, mais oito estudos foram considerados, sendo que apenas um deles foi selecionado. Aplicando-se a pesquisa direta a autores/grupos, mais três estudos foram considerados, sendo dois deles selecionados. Vale destacar que um dos estudos selecionados nessa última etapa levou à exclusão de um outro anteriormente selecionado por meio da aplicação do seguinte critério de exclusão: *O estudo é uma versão mais antiga de um estudo já considerado*. Assim, terminado o processo de seleção, 15 estudos foram selecionados.

Em relação às fontes de pesquisa, no que concerne à busca automática, é importante distinguir entre bases bibliográficas de editoras (tais como *IEEE Xplorer*, *ACM Digital Library* e *Springer Link*) e motores de busca (por exemplo, *Scopus* e *Web of Science*). As bases bibliográficas de editoras contêm basicamente os artigos publicados apenas por elas, enquanto os motores de busca indexam artigos de várias bases bibliográficas. Poder-se-ia pensar que basta, então, fazer a pesquisa em motores abrangendo as principais bases bibliográficas. Contudo, nem sempre os motores de busca cobrem completamente o conteúdo das bases bibliográficas e, muitas vezes, alguns artigos só são retornados a partir das bases bibliográficas das próprias editoras. Esse tem sido o caso, sobretudo, da *IEEE Xplorer* e *ACM Digital Library*. Nesse sentido, recomenda-se usar tanto bases bibliográficas de editoras quanto motores de busca como fontes de pesquisa para a busca automática.

Segundo Dybå et al. [46], o uso das bases bibliográficas da *IEEE Xplorer*⁴² e da *ACM Digital Library*⁴³, combinado com duas bases bibliográficas de motores de busca, por exemplo, *Scopus*⁴⁴ e *Engineering Village*⁴⁵, pode ser suficiente. Ainda que essa diretriz possa ser útil, recomenda-se avaliar se essa combinação é efetivamente suficiente, tendo em vista que o comportamento dos mecanismos de busca tem se alterado nesses últimos anos. Assim, é importante avaliar a cobertura obtida e recomenda-se também avaliar a inclusão das seguintes fontes de dados: *Springer Link*⁴⁶, *ISI Web of Science*⁴⁷ e *Science Direct*⁴⁸. Em todos os MSs realizados mais recentemente, nos quais usamos todas essas fontes, houve casos de estudos que foram retornados somente em uma dessas bases, o que mostra que a diretriz de Dybå et al. [46] não garante a mesma cobertura que a obtida usando todas as bases.

⁴² <http://ieeexplore.ieee.org>

⁴³ <http://dl.acm.org>

⁴⁴ <http://www.scopus.com>

⁴⁵ <http://www.engineeringvillage.com>

⁴⁶ <http://link.springer.com>

⁴⁷ <http://www.webofknowledge.com>

⁴⁸ <http://www.sciencedirect.com>

Deve-se evitar o uso de mecanismos gerais de busca (como o *Google*), pois eles não apresentam o mesmo comportamento ao longo do tempo, comprometendo, assim, a possibilidade de se reproduzir as buscas realizadas. O mesmo ocorre com o *Google Acadêmico* (*Google Scholar*) e, portanto, essa base de dados deve ser evitada como fonte de pesquisa para um MS. Isso não impede que esses mecanismos sejam usados para encontrar estudos retornados na busca. Uma prática bastante comum é fazer a busca automática nas bases bibliográficas das editoras e nos motores de busca e, uma vez selecionados os estudos, procurar seus textos completos utilizando esses mecanismos.

Vale destacar que cada base possui características e limitações próprias e que os motores de busca trabalham de maneiras diferentes. Assim, a *string* de busca deve ser adaptada para rodar adequadamente em cada base [35]. Deve-se observar, dentre outros, se a base considera termos no plural ou se eles devem ser adicionados à *string*, se a base permite realizar buscas considerando apenas partes do texto (por exemplo, título, resumo e palavras-chave) ou se as buscas são feitas sempre considerando o texto completo.

O MS apresentado em [119] foi realizado no contexto de uma tese de doutorado. Havia então uma preocupação em se obter uma ampla cobertura, sobretudo porque, em uma revisão preliminar e informal da literatura, poucos estudos haviam sido encontrados. Dessa forma, optou-se por realizar a busca automática sobre um conjunto amplo de fontes de pesquisa, a saber: *IEEE Xplore*, *ACM Digital Library*, *SpringerLink*, *Scopus*, *Science Direct*, *Engineering Village* e *Web of Science*.

Finalmente, em relação à definição da *string* de busca de um MS, várias abordagens podem ser usadas, tais como [49]: consultar especialistas, melhorar iterativamente a *string*, identificar termos a partir de artigos conhecidos e usar padrões, encyclopédias e tesouros. Vale destacar que essas abordagens não são difíceis de aplicar, não consomem muito tempo e melhoram a qualidade da *string* de busca [49]; portanto, recomenda-se aplicá-las.

Durante a definição da *string* de busca, o foco é a identificação de termos relacionados ao tópico de pesquisa que sejam comumente usados nos estudos primários alvo do MS. Uma boa prática para formular a *string* de busca consiste em agrupar termos relacionados que podem ser considerados sinônimos, concatenando-os em um grupo de termos utilizando-se o conectivo OU (*OR* em inglês). Posteriormente, cada grupo de termos é concatenado com os demais grupos por meio de conectivos E (*AND* em inglês).

Seja o MS apresentado em [119]. Para a formação da *string* de busca, foram considerados dois grupos de termos: um relativo à Gestão do Conhecimento e outro relativo ao Teste de

Software. Para o grupo de termos relacionados à Gestão de Conhecimento, foram identificados os seguintes termos: *knowledge management*, *knowledge reuse*, *knowledge sharing* e *knowledge transfer*; para o grupo de termos relacionados ao teste de software foram identificados os seguintes termos: *software testing* e *software test*, bem como uma combinação dos termos *software project*, *test* e *testing*, resultando na seguinte *string* de busca:

("software testing" OR "software test" OR
("software project" AND (test OR testing)))
AND
("knowledge management" OR "knowledge reuse" OR
"knowledge sharing" OR "knowledge transfer")

Vale ressaltar que os termos da *string* de busca devem estar alinhados ao objetivo e às questões de pesquisa do MS. Essa preocupação pode ser claramente percebida no exemplo discutido acima, no qual a *string* de busca é montada completamente alinhada ao objetivo do mapeamento e às questões de pesquisa, conforme discutido na Questão 4 deste capítulo. Em qualquer caso, a *string* de busca precisa ser avaliada. Especialistas no tema podem avaliar a *string*, mas somente isso é insuficiente. Uma forma comum e efetiva de avaliar a *string* de busca é utilizar um conjunto de estudos primários conhecidos, dito grupo de controle. Nesse caso, a *string* de busca é executada em uma das bases bibliográficas previamente selecionadas (na qual se tenha certeza de que os estudos do grupo de controle estão presentes) e verificase os estudos de controle retornam. Esse procedimento é chamado de teste piloto e serve como um meio de avaliar todo o protocolo do MS (conforme analogamente discutido no Capítulo 2 para o caso de RSs). Para a criação do grupo de controle, recomenda-se a participação de especialistas no tópico do MS, acompanhada de uma revisão informal da literatura.

Para a avaliação da *string* de busca, deve-se considerar sua precisão. Se a *string* de busca retornar um número muito grande de estudos irrelevantes, então vale a pena tentar refiná-la de modo a restringir um pouco mais o número de estudos retornados. A meta é obter uma *string* que retorne um grande número de estudos relevantes, associado ao menor número possível de estudos irrelevantes. Assim, melhorar iterativamente a *string* de busca é uma abordagem que deve ser sempre considerada. De fato, a definição da *string* de busca é um processo iterativo. Na medida que a *string* é testada, em função dos resultados obtidos, novos termos podem ser adicionados à *string*, bem como termos podem ser excluídos por terem se mostrado desnecessários ou pouco eficazes para melhorar a busca.

Questão 7: Quais critérios utilizar para selecionar estudos relevantes em um MS?

Assim como em uma RS, a definição de critérios de seleção é fundamental para garantir a qualidade nos resultados obtidos em um MS. Esses critérios estabelecem características que um estudo deve apresentar para ser considerado relevante para aquele MS (ou seja, os critérios de inclusão) e características que levam à exclusão de estudos que não são relevantes (ou seja, os critérios de exclusão).

Em relação aos critérios de inclusão e exclusão, todas as considerações discutidas na Questão 1 do Capítulo 4 são igualmente aplicáveis aos MSs. Por exemplo, no MS publicado em [119], foram considerados exatamente os critérios de exclusão listados nessa Questão 1, sendo adicionado mais um critério: *O estudo não atender ao critério de inclusão*. Com relação aos critérios de inclusão, somente um critério foi considerado nesse mapeamento: *O estudo discute uma iniciativa de gestão do conhecimento em teste de software*.

Questão 8: Como realizar a avaliação dos estudos para decidir ou não pela sua inclusão?

Conforme discutido na Questão 4 do Capítulo 4, decidir pelos estudos a serem incluídos em um estudo secundário (RS ou MS) envolve julgamento por parte do(s) revisor(es). As quatro configurações apresentadas naquela questão aplicam-se igualmente às RSs e aos MSs. Como se pode notar nessas configurações, o procedimento comum é ter pelo menos dois revisores trabalhando em conjunto [35]. Sendo assim, divergências sobre a inclusão ou não dos estudos podem ocorrer e, por conseguinte, a necessidade de se chegar a um consenso.

No MS apresentado em [119], a doutoranda fez inicialmente a análise de todos os estudos retornados. Todo o conjunto de estudos foi, então, dividido em duas partes e avaliados por seus dois orientadores. Em casos de divergência, procurou-se discutir para chegar a um consenso, às vezes, incluindo os três revisores (a doutoranda e os dois orientadores) na discussão. Casos em que ainda assim persistia a dúvida, os estudos foram incluídos.

Questão 9: É necessário avaliar a qualidade dos estudos selecionados em um MS?

Uma diferença marcante entre RSs e MSs refere-se à avaliação da qualidade dos estudos. Conforme discutido na Questão 3 do Capítulo 4, em RSs, após a seleção final dos estudos, muitas vezes, é realizada a avaliação da qualidade dos estudos selecionados. Em MSs, contudo, a avaliação da qualidade dos estudos selecionados não é necessária, ainda que possa ser útil para garantir que há informações suficientes para a extração de dados [49]. Assim, durante o planejamento de um MS, deve-se decidir se algum critério será usado para avaliar e julgar a qualidade dos estudos selecionados e se a avaliação da qualidade será usada para excluir

estudos abaixo de um certo limiar de qualidade. Além disso, caso decida-se por avaliar a qualidade dos estudos selecionados, é importante encontrar critérios de qualidade que possam ser aplicados de maneira similar nos diferentes tipos de estudos identificados [35].

Petersen et al. [49], ao analisarem 52 MSs, identificaram que apenas 14 deles (~27%) fizeram alguma forma de avaliação da qualidade. Seguindo essa tendência, por exemplo, no MS publicado em [119], decidiu-se não avaliar a qualidade dos estudos selecionados. Além disso, observa-se que uma questão de pesquisa *QP4: Que tipos de pesquisa têm sido feitas?* aborda a questão da qualidade dos estudos.

Questão 10: Como extrair e categorizar dados em um MS?

De acordo com Petersen et al. [120], as atividades em que MSs mais diferem de RSs são a classificação dos estudos e a extração e agregação de dados. A extração de dados realizada em um MS é mais abrangente e tem o objetivo de focar na categorização dos resultados. Em uma RS, a extração de dados é mais detalhada.

Para obter um MS de qualidade, é imprescindível ter um esquema de classificação de estudos confiável e bem definido. Além disso, para aumentar a confiabilidade da classificação realizada em um MS, é muito importante que a extração de dados seja revisada por outro(s) pesquisador(es) e que divergências sejam debatidas, visando atingir um consenso [9] [49].

Para derivar esquemas de classificação, três abordagens principais podem ser adotadas: (i) adotar esquemas existentes; (ii) definir um esquema previamente com base na literatura; ou (iii) deixar os esquemas emergirem dos próprios estudos selecionados, como é o caso das classificações dependentes do tópico do MS. A primeira opção é útil para definir esquemas para questões de pesquisa gerais, que podem ser aplicadas a diferentes MSs. As duas outras abordagens são mais indicadas para derivar esquemas de classificação para questões de pesquisa relacionadas ao tópico específico do MS. Na segunda abordagem (definir o esquema previamente com base na literatura), é muito comum o uso de normas e padrões, tais como normas *ISO/IEC, IEEE* ou corpos de conhecimento como o *SWEBOK* para a ES ou o *PMBOK* para a Gerência de Projetos. Qualquer que seja a abordagem adotada, é sempre aconselhável consultar especialistas na área. Eles podem até mesmo indicar esquemas de classificação existentes [49].

Além disso, Petersen et al. [49] diferenciam entre dois tipos de esquemas de classificação:

- Classificação independente do tópico do MS: referem-se aos esquemas de classificação definidos para questões de pesquisa gerais, presentes em muitos MSs, e que, portanto, não estão diretamente relacionados ao tópico de um MS específico; e
- Classificação dependente do tópico do MS: referem-se aos esquemas de classificação definidos para questões de pesquisa específicas, diretamente relacionadas ao tópico de um particular MS.

No caso dos esquemas de classificação independentes do tópico do MS, é recomendada a utilização dos mesmos esquemas de classificação de (ou similares a) outros MSs. Somente assim, será possível comparar MSs feitos por grupos diferentes [49]. Considerando os aspectos presentes em muitos MSs (ver Questão 4 deste capítulo), é possível adotar esquemas de classificação gerais para tipos de pesquisa, veículos de publicação e métodos de pesquisa [49].

No que se refere aos tipos de pesquisa, a classificação conhecida é a proposta por Wieringa et al. [80] que considera os seguintes tipos de pesquisa:

- Pesquisa de validação: refere-se a estudos validando novas soluções, ainda não implantadas na indústria;
- Pesquisa de avaliação: refere-se a estudos que avaliam soluções na indústria;
- Proposta de solução: diz respeito a estudos que discutem novas soluções (ou revisões de soluções);
- Artigo filosófico: trata de estudos que estruturam um campo de conhecimento de uma nova maneira (por exemplo, uma nova taxonomia);
- Relato de experiência: refere-se a estudos que discutem como algo foi feito na prática; e
- Artigo de opinião: envolve estudos que apresentam a opinião de um pesquisador ou grupo de pesquisadores sobre um tópico de pesquisa específico.

Um mesmo estudo pode ser classificado em mais de uma dessas categorias. Por exemplo, se um estudo apresenta uma nova solução para um problema e discute também uma avaliação em laboratório, ele deve ser classificado como proposta de solução e pesquisa de validação.

Vale destacar que o esquema de classificação por tipos de pesquisa proposto por Wieringa et al. [80] não foi desenvolvido para estudos secundários em geral, mas sim para uso na área de Engenharia de Requisitos. Isso tem feito com que pesquisadores acabem não aplicando largamente esse esquema. É então interessante revisitar esse esquema para ajustá-lo a outros contextos [35]. Recomenda-se também aplicar a tabela de decisão proposta por Petersen et al.

[49], mostrada na Tabela 6.2, que procura auxiliar no uso mais adequado desse esquema. Para o uso dessa tabela, deve-se analisar um estudo primário e identificar em qual das regras (ou seja, colunas de R1 a R6) esse estudo se encaixa, levando-se em consideração as seis condições (apresentadas na primeira coluna). Por exemplo, se um estudo apresentar uma solução sendo utilizada na prática juntamente com uma avaliação empírica, mas não apresentar opinião sobre algo e independentemente se ele apresenta ou não uma nova solução, estrutura conceitual ou experiência do autor, esse estudo se encaixa na regra R1. Tendo a informação da regra em que se encaixa o estudo, é possível decidir a classificação desse estudo considerando uma das seis decisões (listadas na primeira coluna da tabela). Dessa forma, para um estudo que se encaixa na regra R1, pode-se decidir que ele seja do tipo pesquisa de avaliação.

Tabela 6.2: Esquema de classificação de pesquisa. Fonte: [80].

	R1	R2	R3	R4	R5	R6
Condições						
<i>Usado na prática</i>	V	•	V	F	F	F
<i>Nova solução</i>	•	V	F	•	F	F
<i>Avaliação empírica</i>	V	F	F	V	F	F
<i>Estrutura conceitual</i>	•	•	•	•	V	V
<i>Opinião sobre algo</i>	F	F	F	F	F	V
<i>Experiência do autor</i>	•	•	•	•	F	F
Decisões						
<i>Pesquisa de avaliação</i>	✓	•	•	•	•	•
<i>Proposta de solução</i>	•	✓	•	•	•	
<i>Pesquisa de validação</i>	•	•	•	✓	•	•
<i>Artigos filosóficos</i>	•	•	•	•	✓	•
<i>Artigos de opinião</i>	•	•	•	•	•	✓
<i>Relato de experiência</i>	•	•	✓	•	•	•
Legenda: V = Verdadeiro; F = Falso; • = irrelevante ou não aplicado; R1-R6 = referentes às regras aplicadas						

No que se refere aos tipos de veículos de publicação, uma classificação bastante adotada considera três categorias básicas: periódicos, conferências (incluindo simpósios) e *workshops* (eventos menores, normalmente colocalizados com ou realizados como eventos satélites de conferências).

Por fim, no tocante à classificação por métodos de pesquisa, Petersen et al. [49] propõem um esquema de classificação baseado em alguns trabalhos, dentre eles [8] e [124]. Esse esquema inclui os seguintes métodos de pesquisa frequentemente aplicados na ES: *survey*, estudo de caso, experimento controlado, pesquisa ação, etnografia, simulação, prototipagem e análise e prova matemática.

Petersen et al. [49] ainda argumentam que a classificação por métodos de pesquisa tem de estar consistente com a classificação por tipo de pesquisa, em especial no que se refere à distinção entre pesquisa de avaliação e pesquisa de validação. Pesquisa ação e etnografia só se aplicam a situações envolvendo profissionais e, portanto, o tipo de pesquisa relacionado deve ser obrigatoriamente a pesquisa de avaliação. Por outro lado, prototipagem, simulação e análise e prova matemática são métodos de pesquisa usados essencialmente em pesquisa de validação. Note que certos métodos de pesquisa podem estar relacionados a ambas as categorias de tipos de pesquisa. Esse é o caso de experimentos, surveys e estudos de caso que, se realizados com estudantes, são classificados como pesquisa de validação, enquanto se realizados com profissionais, são classificados como pesquisa de avaliação. A Figura 6.1 ilustra essa relação.

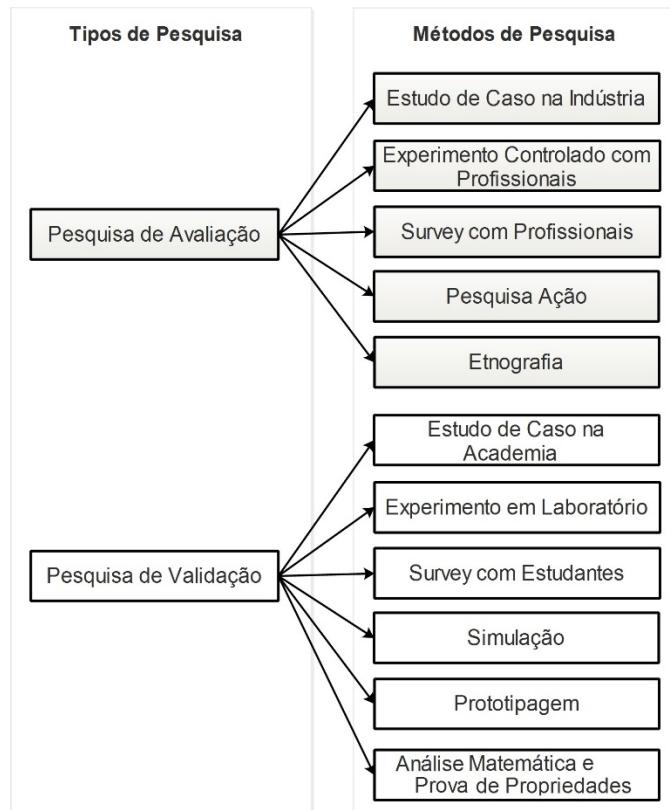


Figura 6.1: Relação entre tipos de pesquisa e métodos de pesquisa (adaptado de [49]).

No MS apresentado em [119], para algumas questões de pesquisa (a saber: QP2, QP3, QP5, QP6, QP8 e QP9, listadas na Questão 5 deste capítulo), os esquemas de classificação emergiram dos próprios estudos selecionados; para outras duas (a saber: QP1 e QP4), foram adotados esquemas de classificação existentes (respectivamente, os esquemas por veículos de publicação e tipos de pesquisa). Finalmente para uma delas (QP7), o esquema de classificação foi definido com base na literatura (foi usada a classificação proposta por Nonaka e Takeuchi

[125] para tipos de itens de conhecimento, que considera dois tipos de itens de conhecimento: conhecimento tácito e conhecimento explícito), sendo que esse esquema foi posteriormente refinado com base nos próprios estudos selecionados.

Questão 11: Como sintetizar e apresentar os resultados de um MS?

Conforme discutido anteriormente, um MS foca a categorização dos estudos. Assim, é muito comum que os resultados sejam sintetizados e apresentados na forma de tabelas e gráficos, com destaque para gráficos de barras, gráficos de pizza e gráficos de bolhas.

Gráficos de barras e gráficos de pizza permitem representar adequadamente o número de estudos classificados em cada categoria. Tabelas também servem a esse propósito, sendo possível indicar, ainda, quais estudos foram classificados em cada categoria. Contudo, às vezes, é interessante combinar informações sobre diferentes categorias. Para agregar dados de diferentes categorias e extraír informações mais complexas, derivadas do cruzamento de duas categorias, gráficos de bolha são mais indicados. Um gráfico de bolha é uma variação de um gráfico de dispersão em que os pontos de dados são substituídos por bolhas e uma dimensão adicional dos dados é representada pelo tamanho das bolhas, possibilitando representar informações em três dimensões. Quando usados em MSs, o tamanho da bolha é proporcional ao número de estudos que se enquadram em um par de categorias (por exemplo, subtópico de pesquisa combinado com tipo de pesquisa) [49]. Exemplos de MSs que adotaram esse tipo de representação podem ser encontrados em [126] [127]. Outras formas de visualização usadas em MSs incluem mapas de calor, gráficos de linhas e diagramas de Venn [49].

No MS apresentado em [119], foram apresentados um gráfico de barras (para mostrar a distribuição dos estudos ao longo do tempo) e diversas tabelas, permitindo não apenas apresentar o número (e o percentual) de estudos classificados em cada categoria, mas também quais são esses estudos. Vale destacar que essa forma de apresentação foi possível em função da pequena quantidade de estudos selecionados, no caso apenas 15 estudos.

Questão 12: Quais são as ameaças à validade de um MS?

Por melhor que um MS tenha sido planejado e conduzido, sempre haverá ameaças à validade nos resultados do mapeamento. Nesse sentido, é fundamental identificar e explicitar essas ameaças buscando deixar claro para o leitor o que pode ter comprometido os resultados, dando a ele uma base para definir se esse mapeamento lhe será útil ou não. Possíveis ameaças à validade de um MS incluem:

- O conjunto de estudos primários retornados não corresponde ao conjunto efetivamente existente na literatura sobre o tópico daquele mapeamento;
- Possível avaliação tendenciosa por parte dos pesquisadores envolvidos na seleção dos estudos;

- Capacidade de generalizar os resultados do mapeamento. Essa ameaça é mais forte quando poucos estudos são selecionados e não há evidências suficientes para se generalizar os resultados; e
- Confiabilidade das conclusões em relação aos dados coletados, por exemplo, devido a uma possível tendência dos pesquisadores envolvidos na interpretação dos dados.

No MS apresentado em [119], foram apresentadas as ameaças à validade na forma de limitações, bem como as ações que foram tomadas para minimizar essas ameaças.

Questão 13: Como reportar um MS?

O processo de realização de MSs encerra-se com a fase de publicação dos resultados do mapeamento, quando os resultados devem ser reportados em um relatório ou artigo. Dado que, em geral, MSs em si têm grande valor para a comunidade científica da área abordada, recomenda-se a sua publicação na forma de um artigo. Tanto quanto possível, deve-se procurar reportar um MS seguindo a mesma estrutura e estilo de publicações já feitas apresentando MSs. Isso torna mais fácil avaliar e comparar estudos dessa natureza [49]. Uma estrutura de artigo indicada inclui as seguintes seções:

- Introdução: deve prover uma visão geral inicial sobre o tópico sendo investigado e também uma motivação para a importância desse tópico, de modo a justificar a necessidade de se realizar um MS sobre esse tópico. A introdução deve ainda destacar a utilidade desse MS, podendo também prover uma visão geral de como ele foi conduzido;
- Fundamentação teórica: deve prover uma fundamentação teórica do tópico alvo do MS. Trabalhos da literatura que tenham sido usados como base para a formulação de questões de pesquisa ou esquemas de classificação devem ser abordados. Trabalhos relacionados (ou seja, outros estudos de revisões da literatura conduzidas de forma sistemática ou não), se existentes, devem se brevemente comentados, procurando justificar porque eles não são suficientes para prover uma visão geral e abrangente do tópico alvo do mapeamento;
- Método de pesquisa: nessa seção devem ser apresentadas as principais informações relacionadas ao método de pesquisa aplicado (no caso, MS). Devem ser claramente apresentados: (i) objetivo do MS; (ii) questões de pesquisa; (iii) aspectos básicos relacionados à busca e seleção, incluindo: *string* de busca, fontes pesquisadas, critérios de inclusão e de exclusão, bem como os critérios de qualidade (quando pertinente); (iv) informações sobre como os dados foram coletados e armazenados; (v)

informações sobre como a avaliação foi realizada; e (vi) ameaças à validade do estudo (alternativamente, essa informação pode ser apresentada no final da seção de resultados);

- Resultados: essa seção deve iniciar dando uma visão geral do processo de seleção e dos estudos retornados e selecionados nas duas etapas do MS. É fundamental prover uma lista dos estudos selecionados. A seguir, devem ser apresentados os esquemas de classificação adotados e os resultados estruturados por questão de pesquisa. Recomenda-se apresentar apenas os resultados;
- Discussão: deve ser apresentada uma discussão com base na interpretação e análise dos resultados obtidos e apresentados na seção anterior. Tendências e lacunas de pesquisa devem ser também apontadas; e
- Conclusões: nessa seção deve ser apresentado um resumo do MS e suas conclusões, bem como os trabalhos futuros.

Vale destacar que uma publicação reportando os resultados de um MS deve listar todos os estudos selecionados. Um mapeamento que não apresenta essa lista de estudos tem, geralmente, sua utilidade diminuída para outros pesquisadores. Uma razão comum para que isso não seja feito é a limitação imposta pelas conferências (e até alguns periódicos) para o número de páginas da publicação. Assim, recomenda-se escolher um veículo de publicação que não imponha muitas restrições. Uma alternativa interessante para evitar esse problema consiste em incluir detalhes sobre todos os estudos primários selecionados e suas classificações em um texto publicado na web [121], como fizeram Neto et al. [128].

Questão 14: Quais são os principais problemas enfrentados durante a realização de um MS?

Ainda que haja vários benefícios em se realizar um MS, também há problemas e desafios. Em primeiro lugar, deve-se considerar que, dependendo do volume de pesquisas já realizadas no tópico sendo pesquisado, a realização de um MS pode consumir muito tempo. Além disso, dois problemas recorrentes em MSs são a seleção e a classificação. Ambos são um desafio, sobretudo para estudantes ou pesquisadores pouco familiarizados com o tópico de pesquisa em questão. Eles podem não estar familiarizados totalmente com o tópico e pode lhes faltar um entendimento mais sólido da terminologia empregada nos estudos. Mesmo pesquisadores experientes podem fazer avaliações e classificações substancialmente diferentes para um mesmo trabalho. Normalmente, esse problema pode ser contornado com a participação de vários pesquisadores. Contudo, nem sempre isso é possível. No caso de projetos de mestrado

ou doutorado, é natural que o orientador atue na avaliação, tanto da seleção quanto da classificação de estudos.

Um desafio ainda maior é avaliar a qualidade dos estudos selecionados. Ainda que essa não seja obrigatória para MSs, quando realizada, gera dificuldades, tais como na definição dos critérios de qualidade e na avaliação dos estudos primários em si. Esse desafio é ainda maior devido ao fato de muitos estudos, por serem apresentados em tamanho relativamente reduzido, apresentarem poucas informações sobre o trabalho realizado [122].

Considerações finais

Um MS é um estudo secundário que tem como objetivo identificar e classificar a pesquisa relacionada a um tópico mais amplo de pesquisa [7]. Há muitas semelhanças e diferenças entre MSs e RSs, sendo que ambos possibilitam a identificação e agregação das evidências disponíveis para responder às questões de pesquisa, auxiliam na identificação de lacunas de pesquisa e compartilham um processo comum para a sua execução. Contudo, há também diferenças.

Em resumo, os principais pontos de aprendizagem deste capítulo foram:

- MS e RS apresentam muitas semelhanças, mas também diferenças. As principais diferenças entre MS e RS residem no objetivo, escopo e nos procedimentos de análise de estudos;
- Há vários benefícios em se realizar um MS, inclusive em conjunto com uma RS. Contudo, há também problemas e desafios;
- As fases do processo de MS são as mesmas do processo de RS: planejamento, condução e publicação dos resultados. Contudo, há diretrizes específicas para a realização dessas fases, quando realizadas no contexto de um MS ou de uma RS;
- As questões de pesquisa de um MS são questões mais gerais (ou então, mais abrangentes) do que as questões de pesquisa de uma RS;
- Para que um MS forneça uma visão ampla de um tópico de pesquisa, deve-se necessariamente realizar a busca automática. A busca manual, *snowballing* e pesquisa direta a pesquisadores/grupos de pesquisa são estratégias complementares a serem usadas, quando pertinente, em conjunto com a busca automática;
- Os termos da *string* de busca devem estar alinhados ao objetivo e às questões de pesquisa do MS, sendo que ela deve ser avaliada por meio de testes piloto;

- Critérios de inclusão e exclusão devem ser definidos, enquanto que os critérios de qualidade podem não ser necessários, visto que a avaliação da qualidade dos estudos primários não é obrigatória;
- Para obter um MS de qualidade é imprescindível ter um esquema de classificação confiável e bem definido;
- É útil apresentar os resultados de um MS na forma de tabelas e gráficos, com destaque para gráficos de barras, gráficos de pizza e gráficos de bolhas;
- É fundamental identificar ameaças à validade do mapeamento para deixar claro para o leitor o que pode ter comprometido os resultados; e
- MSs em si têm grande valor para a comunidade científica relacionada ao tópico de pesquisa e, portanto, esses mapeamentos deveriam ser publicados. Tanto quanto possível, deve-se reportar um MS seguindo a estrutura e estilo de publicações já realizadas apresentando MSs.

7. Considerações finais

Katia Romero Felizardo
Elisa Yumi Nakagawa
Sandra Camargo Pinto Ferraz Fabbri
Fabiano Cutigi Ferrari

O objetivo deste livro foi proporcionar ao leitor um entendimento sobre o processo de RS, buscando sumarizar as pesquisas sendo conduzidas na área de RS, bem como compartilhar a nossa experiência e prover orientações aos pesquisadores que pretendem conduzir suas revisões. Neste livro foram abordadas as três fases que compõem o processo de RS (planejamento, condução e publicação de resultados), juntamente com as suas diversas atividades. Vale destacar que a comunidade de ES que tem largamente conduzido RSs nos últimos anos reconhece a natureza iterativa desse processo e, dessa forma, os artefatos resultantes podem ser refinados à medida que as iterações ocorrem.

Pode-se afirmar ainda que a RS, juntamente com o MS, têm sido compreendidos pela ES como sendo métodos de pesquisa importantes, senão essenciais, para pesquisadores que visam conhecer e caracterizar o estado da arte da pesquisa no que concerne aos diversos tópicos de pesquisa em ES. Nessa perspectiva, observa-se um número crescente de periódicos de alto impacto e eventos científicos de primeira linha da área de ES que têm publicado estudos secundários, tais como o *Journal of Systems and Software*; o *Journal of Information and Software Technology*; o *IEEE Transactions on Software Engineering*; o *Journal Software: Practice and Experience*; e o *Empirical Software Engineering*. Salientamos também que nos últimos anos o amadurecimento da pesquisa sobre RS é perceptível. Além disso, é inegável a relevância e as vantagens decorrentes de sua execução. Assim, os pesquisadores têm em mãos uma maneira sistemática de organizar o conhecimento existente, identificar mais pontualmente as lacunas de pesquisa e relacionar o conhecimento dos diversos tópicos. Consequentemente, por meio de RSs e MSs é possível prover uma base sólida de conhecimento a partir da qual a própria área de ES tem se respaldado para evoluir seu estado da arte, o que contribui para seu principal objetivo que é o desenvolvimento de sistemas de software de qualidade.

Em uma outra perspectiva, vale destacar que a execução de RSs não é uma tarefa trivial, mesmo para pesquisadores que tenham experiência tanto em RS quanto no tópico pesquisado. Há ainda diversos desafios e limitações inerentemente associadas à sua execução. Em particular, há um consenso de que a seleção de estudos primários, a avaliação da qualidade desses estudos e a extração de dados são atividades que impõem dificuldade e

consumem um tempo significativo na execução do processo. Além disso, a busca por estudos primários é um desafio, pois a qualidade da busca interfere diretamente na qualidade dos estudos identificados e nos resultados. Salienta-se que além da perspicácia e de um conhecimento básico necessários ao pesquisador, as bases de busca (bases bibliográficas e motores de busca) não estão preparadas adequadamente para dar subsídios à execução de estudos secundários. Observa-se ainda que o conhecimento (registrado e publicado por meio de estudos) encontra-se hoje espalhado em diversas fontes acessíveis de maneiras bastante distintas.

Ao longo da execução de RSs, há ainda a necessidade de lidar com diversos *trade-offs*. Somente a critério de exemplo, durante a atividade da busca e identificação de estudos, uma busca muito abrangente pode requerer, posteriormente, um esforço considerável para a seleção dos estudos que são realmente relevantes; por outro lado, uma busca restritiva pode interferir na acurácia e qualidade dos resultados da revisão. Nesse cenário, para dar apoio a essas e outras atividades, suporte computacional ao processo de RS e MS torna-se essencial. Apesar de já existirem iniciativas nesse sentido, ainda há muito a evoluir.

Da perspectiva dos pesquisadores da área de RS, recomendamos aos autores de estudos primários que deem atenção para a preparação dos resumos e seleção das palavras-chave de seus estudos, pois isso afeta a chance deles serem encontrados nas bases de busca e de serem selecionados pelos pesquisadores, caso sejam pertinentes ao tópico investigado na revisão que está sendo conduzida. Ainda referente à busca e identificação de estudos, vale salientar que a área de ES, apesar do seu quase meio século de existência, ainda carece de uma padronização geral nos termos da área, além, muitas vezes, da falta de adoção de padrões já existentes pela própria comunidade. Isso acarreta um esforço adicional na identificação de estudos relevantes para um dado tópico de pesquisa. Além disso, salientamos aos autores de estudos primários a importância da condução de estudos experimentais que promovam evidências concretas sobre o tópico investigado nos seus respectivos trabalhos.

Quanto à atividade de síntese de dados, salientamos a importância, o conhecimento necessário e a aplicação dos métodos de síntese existentes, sejam eles quantitativos, qualitativos ou mistos. Um outro desafio é utilizar, na prática, os resultados de RSs nas decisões tomadas no desenvolvimento de software na indústria, uma vez que isso não tem sido claramente observado, distanciando-se de um dos propósitos originais de se conduzir RSs. No entanto, observamos que atualmente os estudos secundários (RSs e MSs) têm sido importantes e servido como base para a condução da pesquisa em ES. Muitas vezes, RSs e MSs

têm sido executados no contexto de projetos de pesquisa e/ou projetos de doutorado e publicados em veículos de difusão científica.

Por fim, os autores deste livro agradecem novamente aos colaboradores e também aos leitores e os convidam a enviar perguntas, comentários e questionamentos para uma possível nova edição deste livro. Por favor, contactar a Profa. Katia Romero Felizardo através do email *katiascannavino@utfpr.edu.br*.

8. Referências Bibliográficas

- [1] B. A. Kitchenham, T. Dybå, and M. Jørgensen, "Evidence-based software engineering," in *26th International Conference on Software Engineering (ICSE 2004)*, 2004, pp. 273–281.
- [2] T. Dybå, "Evidence-based software engineering for practitioners," *IEEE Softw.*, vol. 22, no. 1, pp. 58–65, 2005.
- [3] V. R. Basili and M. V Zelkowitz, "Empirical studies to build a science of computer science," *Commun. ACM*, vol. 50, no. 11, pp. 33–37, 2007.
- [4] B. A. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.
- [5] F. Q. da Silva, A. L. Santos, S. Soares, A. C. C. Franca, C. V Monteiro, and F. F. Maciel, "Six years of systematic literature reviews in software engineering: an updated tertiary study," *Inf. Softw. Technol.*, vol. 53, no. 9, pp. 899–913, 2011.
- [6] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos, "Systematic review in software engineering," 2005.
- [7] B. A. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering -- version 2.3," Keele/Staffs-UK and Durham-UK, 2007.
- [8] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering: an introduction*. Springer, 2012.
- [9] P. O. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, 2007.
- [10] B. A. Kitchenham, "Procedures for performing systematic reviews," 2004.
- [11] S. N. Mafra and G. H. Travassos, "Estudos primários e secundários apoiando a busca por evidência em engenharia de software," 2006.
- [12] D. L. Sackett, W. M. C. Rosenberg, J. A. M. Gray, R. B. Haynes, and W. S. Richardson, "Evidence based medicine: what it is and what it isn't," *Br. Med. J.*, vol. 1, no. 316, pp. 71–72, 1996.

- [13] D. L. Sackett, S. E. Straus, W. S. Richardson, W. Rosenberg, and R. B. Haynes, *Evidence-based medicine: how to practice and teach {EBM}*. Churchill Livingstone: Edinburgh, 2000.
- [14] H. Zhang and M. A. Babar, "Systematic reviews in software engineering: An empirical investigation," *Inf. Softw. Technol.*, vol. 55, no. 7, pp. 1341–1354, 2013.
- [15] B. A. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering - A tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010.
- [16] J. M. Verner, O. P. Brereton, B. A. Kitchenham, M. Turner, and M. Niazi, "Systematic literature reviews in global software development: A tertiary study," in *16th International Conference on Evaluation & Assessment in Software Engineering (EASE 2012)*, 2012, pp. 2–11.
- [17] S. Imtiaz, M. Bano, N. Ikram, and M. Niazi, "A tertiary study: experiences of conducting systematic literature reviews in software engineering," in *17th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2013)*, 2013, pp. 177–182.
- [18] S. C. P. F. Fabbri, K. R. Felizardo, F. C. Ferrari, E. C. M. Hernandes, F. R. Octaviano, E. Y. Nakagawa, and J. C. Maldonado, "Externalising tacit knowledge of the systematic review process," *IET Softw.*, vol. 7, no. 6, pp. 298–307, 2013.
- [19] M. Riaz, N. Sulayman M. Salleh, and E. Mendes, "Experiences Conducting Systematic Reviews from Novices' Perspective," in *13rd International Conference on Evaluation and Assessment in Software Engineering (EASE' 10)*, 2010, pp. 1–10.
- [20] T. e I. E. D. de C. e T. da Saúde Secretaria de Ciência, "Diretrizes metodológicas: elaboração de revisão sistemática e metanálise de ensaios clínicos randomizados," 2012.
- [21] J. C. Carver, E. Hassler, E. Hernandes, and N. A. Kraft, "Identifying barriers to the systematic literature review process," in *7th International Symposium on Empirical Software Engineering and Measurement (ESEM 2013)*, 2013, pp. 203–212.
- [22] J. C. A. Biolchini, P. G. Mian, A. C. C. Natali, T. U. Conte, and G. H. Travassos, "Scientific research ontology to support systematic review in software engineering," *Adv. Eng. Informatics*, vol. 21, no. 2, pp. 133–151, 2007.
- [23] M. A. Babar and H. Zhang, "Systematic literature reviews in software engineering:

- preliminary results from interviews with researchers,” in *3rd International Symposium on Empirical Software Engineering and Measurement (ESEM 2009)*, 2009, pp. 346–355.
- [24] C. Marshall and P. O. Brereton, “Tools to support systematic literature reviews in software engineering: a mapping study,” in *7th International Symposium on Empirical Software Engineering and Measurement (ESEM 2013)*, 2013, pp. 296–299.
 - [25] C. Marshall, P. O. Brereton, and B. A. Kitchenham, “Tools to support systematic reviews in software engineering: a feature analysis,” in *18th International Conference on Evaluation Assessment in Software Engineering (EASE 2014)*, 2014, pp. 13:1–13:10.
 - [26] C. Marshall and P. O. Brereton, “Systematic review toolbox: a catalogue of tools to support systematic reviews,” in *19th International Conference on Evaluation Assessment in Software Engineering (EASE 2015)*, 2015, pp. 23:1–23:6.
 - [27] F. Tomassetti, G. Rizzo, A. Vetro, L. Ardito, M. Torchiano, and M. Morisio, “Linked data approach for selection process automation in systematic reviews,” in *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, 2011, pp. 31–35.
 - [28] K. R. Felizardo, G. F. Andery, F. V Paulovich, R. Minghim, and J. C. Maldonado, “A visual analysis approach to validate the selection review of primary studies in systematic reviews,” *Inf. Softw. Technol.*, vol. 54, no. 10, pp. 1079–1091, 2012.
 - [29] J. A. S. Torres, D. S. Cruzes, and L. N. Salvador, “Automatic results identification in software engineering papers. {I}s it possible?,” in *12th International Conference on Computational Science and Its Applications (ICCSA 2012)*, 2012, pp. 108–112.
 - [30] K. R. Felizardo, S. G. MacDonell, E. Mendes, and J. C. Maldonado, “A Systematic Mapping on the use of Visual Data Mining to Support the Conduct of Systematic Literature Reviews,” in *Journal of Software (JSW)*, 2011, pp. 1–10.
 - [31] D. Bowes, T. Hall, and S. Beecham, “{SLuRp}: a tool to help large complex systematic literature reviews deliver valid and rigorous results,” in *2nd International Workshop on Evidential Assessment of Software Technologies (EAST 2012)*, 2012, pp. 33–36.
 - [32] E. C. M. Hernandes, A. B. Zamboni, A. D. Thommazo, and S. C. P. F. Fabbri, “Using {GQM} and {TAM} to evaluate {StArt} -- a tool that supports systematic review,” *CLEI Eletronic J.*, vol. 15, no. 1, pp. 1–13, 2012.
 - [33] A. M. Fernandez-Saez, M. G. Bocco, and F. P. Romero, “SLR-TOOL - A Tool for

- Performing Systematic Literature Reviews," in *4th International Conference on Software and (ICSOFT' 10)*, 2010, pp. 1–10.
- [34] B. S. Barn, F. Raimondi, L. Athappian, and T. Clark, "{SLRT}ool: a tool to support collaborative systematic literature reviews," in *16th International Conference on Enterprise Information Systems (ICEIS 2014)*, 2014, pp. 440–447.
 - [35] C. Wohlin, P. Runeson, M. Host, P. A. M. Silveira Neto, E. Engstrom, I. C. Machado, and E. S. Almeida, "On the reliability of mapping studies in software engineering," *J. Syst. Softw.*, vol. 86, no. 10, pp. 2594–2610, 2013.
 - [36] S. MacDonell, M. Shepperd, B. A. Kitchenham, and E. M. I. C. Society, "How Reliable Are Systematic Reviews in Empirical Software Engineering?," *IEEE Trans. Softw. Eng.*, vol. 36, no. 5, pp. 676–687, 2010.
 - [37] B. A. Kitchenham, P. O. Brereton, Z. Li, D. Budgen, and Burn, "Repeatability of systematic literature reviews," in *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, 2011, pp. 46–55.
 - [38] P. O. Brereton, "A study of computing Undergraduates undertaking a systematic literature review," *IEEE Trans. Educ.*, vol. 54, no. 4, pp. 558–563, 2011.
 - [39] S. G. MacDonell and M. J. Shepperd, "Comparing Local and Global Software Effort Estimation Models -- Reflections on a Systematic Review," in *1st International Symposium on Empirical Software Engineering and Measurement (ESEM' 07)*, 2007, pp. 401–409.
 - [40] M. Jørgensen and M. Shepperd, "A systematic review of software development cost estimation studies," *IEEE Trans. Softw. Eng.*, vol. 33, no. 1, pp. 33–53, 2007.
 - [41] T. B. C. Arias, P. van der Spek, and P. Avgeriou, "A practice-driven systematic review of dependency analysis solutions," *Empir. Softw. Eng.*, vol. 16, no. 5, pp. 544–586, 2011.
 - [42] M. Galster, D. Weyns, D. Tofan, B. Michalik, and P. Avgeriou, "Variability in software systems -- a systematic literature review," *IEEE Trans. Softw. Eng.*, vol. 40, no. 3, pp. 282–306, 2014.
 - [43] E. Mendes, M. Kalinowski, D. Martins, F. Ferrucci, and F. Sarro, "Cross- vs. within-company cost estimation studies revisited: an extended systematic review," in *18th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2014)*, 2014, pp. 1–10.

- [44] M. Pai, M. McCulloch, J. D. Gorman, N. Pai, W. Enanoria, G. Kennedy, P. Tharyan, and J. M. Colford, "Systematic reviews and meta-analyses: an illustrated, step-by-step guide," *Natl. Med. J. India*, vol. 17, no. 2, pp. 89–95, 2004.
- [45] L. B. R. Oliveira, F. S. Osório, and E. Y. Nakagawa, "A systematic review on service-oriented robotic systems development," São Carlos: Biblioteca Prof. Achilles Bassi, 2012.
- [46] T. Dybå, T. Dingsøyr, and G. K. Hanssen, "Applying systematic reviews to diverse study types: an experience report," in *1st International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, 2007, pp. 225–234.
- [47] M. Staples and M. Niazi, "Experiences using systematic review guidelines," *J. Syst. Softw.*, vol. 80, no. 9, pp. 1425–1437, 2007.
- [48] K. S. Khan, R. Kunz, J. Kleijnen, and G. Antes, "Systematic reviews to support evidence-based medicine: how to review and apply findings of healthcare research," 2003.
- [49] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, pp. 1–18, 2015.
- [50] H. Zhang and A. B. Muhammad, "On Searching Relevant Studies in Software Engineering," in *13rd International Conference on Evaluation and Assessment in Software Engineering (EASE' 10)*, 2010, pp. 1–10.
- [51] O. Dieste, A. Grimán, and N. Juristo, "Developing search strategies for detecting relevant experiments," *Empir. Softw. Eng.*, vol. 14, no. 5, pp. 513–539, 2009.
- [52] J. A. M. Santos, A. R. Santos, and M. Mendonça, "Investigating bias in the search phase of software engineering secondary studies," in *18th Ibero-American Conference on Software Engineering (ESELAW 2015)*, 2015, pp. 488–501.
- [53] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014)*, 2014, pp. 1–10.
- [54] A. Booth, "Unpacking your literature search toolbox: on search styles and tactics," *Heal. Inf. Libr. J.*, vol. 25, no. 4, pp. 313–317, 2008.
- [55] K. R. Stevens, "Systematic reviews: the heart of evidence-based practice," *AACN Adv.*

Crit. Care, vol. 12, no. 4, pp. 529–538, 2001.

- [56] C. M. C. Santos, C. A. M. Pimenta, and M. R. C. Nobre, “A estrat{é}gia {PICO} para a constru{ç}{ão} da pergunta de pesquisa e busca de evid{ê}ncias,” *Rev. Lat. Am. Enferm.*, vol. 15, no. 3, pp. 508–511, 2007.
- [57] IEEE, “{IEEE} standard glossary of software engineering terminology,” *IEEE Std 610.12-1990*, 1990.
- [58] L. Chen, M. A. Babar, and H. Zhang, “Towards an evidence-based understanding of electronic data sources,” in *14th International Conference on Evaluation and Assessment in Software Engineering (EASE 2010)*, 2010, pp. 135–138.
- [59] H. Zhang, M. A. Babar, and P. Tell, “Identifying relevant studies in software engineering,” *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 625–637, 2011.
- [60] L. B. R. Oliveira, F. S. Osório, and E. Y. Nakagawa, “An investigation into the development of service-oriented robotic systems,” in *28th Annual ACM Symposium on Applied Computing (SAC 2013)*, 2013, pp. 223–228.
- [61] J. DeLuca, M. Mullins, L. Cynthia, N. Crepaz, L. Kay, and S. Thadiparthi, “Developing a comprehensive search strategy for evidence based systematic reviews,” *Evid. Based Libr. Inf. Pract.*, vol. 3, no. 1, pp. 3–32, 2008.
- [62] M. K. Choong, F. Galgani, A. G. Dunn, and G. Tsafnat, “Automatic evidence retrieval for systematic reviews,” *J. Med. Internet Res.*, vol. 16, no. 10, 2014.
- [63] B. A. Kitchenham and P. O. Brereton, “A systematic review of systematic review process research in software engineering,” *Inf. Softw. Technol.*, vol. 55, no. 12, pp. 2049–2075, 2013.
- [64] V. Alberani, P. C. Pietrangeli, and A. M. Mazza, “The use of grey literature in health sciences: a preliminary survey.,” *Bull. Med. Libr. Assoc.*, vol. 78, no. 4, p. 358, 1990.
- [65] R. Helms-Park, P. Radia, and P. Stapleton, “A preliminary assessment of {G}oogle {S}cholar as a source of {EAP} students’ research materials,” *Internet High. Educ.*, vol. 10, no. 1, pp. 65–76, 2007.
- [66] D. Badampudi, C. Wohlin, and K. Petersen, “Experiences from using snowballing and database searches in systematic literature studies,” in *19th International Conference on Evaluation and Assessment in Software Engineering (EASE 2015)*, 2015, pp. 17:1–17:10.

- [67] F. C. Ferrari and J. C. Maldonado, "Experimenting with a multi-iteration systematic review in software engineering," in *5th Experimental Software Engineering Latin American Workshop (ESELAW 2008)*, 2008, pp. 1–10.
- [68] S. Cruz, F. Q. B. da Silva, and L. F. Capretz, "Forty years of research on personality in software engineering: A mapping study," *Comput. Human Behav.*, vol. 46, pp. 94–113, May 2015.
- [69] M. C. Conry, N. Humphries, K. Morgan, Y. McGowan, A. Montgomery, K. Vedhara, E. Panagopoulou, and H. McGee, "A 10 year (2000–2010) systematic review of interventions to improve quality of care in hospitals," *BMC Health Serv. Res.*, vol. 12, no. 275, pp. 1–16, 2012.
- [70] M. Bano and D. Zowghi, "A systematic review on the relationship between user involvement and system success," *Inf. Softw. Technol.*, vol. 58, pp. 148–169, 2015.
- [71] R. W. Lam and S. H. Kennedy, "Using metaanalysis to evaluate evidence: practical tips and traps," *Can. J. Psychiatry*, vol. 50, no. 3, pp. 167–174, 2005.
- [72] K. R. Felizardo, M. Riaz, S. Muhammad, E. Mendes, S. G. MacDonell, and J. C. Maldonado, "Analysing the use of graphs to represent the results of Systematic Reviews in Software Engineering," in *25th Brazilian Symposium on Software Engineering (SBES' 11)*, 2011, pp. 1–10.
- [73] B. A. Kitchenham, O. P. Brereton, S. Owen, J. Butcher, and C. Jefferies, "Length and readability of structured software engineering abstracts.," *IET Softw.*, vol. 2, no. 1, pp. 37–45, 2008.
- [74] H. Zhang and A. B. Muhammad, "An Empirical Investigation of Systematic Reviews in Software Engineering," in *5th International Symposium on Empirical Software Engineering and Measurement (ESEM' 11)*, 2011, pp. 1–10.
- [75] T. Hall, S. Beecham, D. Bowes, D. Gray, and S. Counsell, "A systematic literature review on fault prediction performance in software engineering," *IEEE Trans. Softw. Eng.*, vol. 38, no. 6, pp. 1276–1304, 2012.
- [76] M. B. Blake, S. L. Remy, Y. Wei, and A. M. Howard, "Robots on the web," *Robot. Autom. Mag.*, vol. 18, no. 2, pp. 33–43, 2011.
- [77] D. Doran and S. S. Gokhale, "Searching for heavy tails in web robot traffic," in *17th International Conference on the Quantitative Evaluation of Systems (QEST 2010)*, 2010,

pp. 282–291.

- [78] W. T. Tsai, X. Sun, Q. Huang, and H. Karatza, “An ontology-based collaborative service-oriented simulation framework with {Microsoft Robotics Studio},” *Simul. Model. Pract. Theory*, vol. 16, no. 9, pp. 1392–1414, 2008.
- [79] P. Zhang, R. Lu, and J. Zhang, “Reliability modeling and analysis of service-oriented robot management system,” in *IEEE Asia-Pacific Conference on Services Computing (APSCC 2006)*, 2006, pp. 621–624.
- [80] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, “Requirements engineering paper classification and evaluation criteria: a proposal and a discussion,” *Requir. Eng.*, vol. 11, no. 1, pp. 102–107, 2005.
- [81] P. Edwards, M. Clarke, C. DiGuiseppi, S. Pratap, I. Roberts, and R. Wentz, “Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records,” *Stat. Med. J.*, vol. 21, no. 11, pp. 1635–1640, 2002.
- [82] O. Dieste and N. Juristo, “Systematic review and aggregation of empirical studies on elicitation techniques,” *IEEE Trans. Softw. Eng.*, vol. 37, no. 2, pp. 283–304, 2011.
- [83] S. Siegel and N. J. Castellan Jr., *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, 1988.
- [84] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
- [85] A. Osborne-OHagan, G. Coleman, and R. V OConnor, “Software development processes for games: a systematic literature review,” in *21st European Conference on Systems, Software and Services Process Improvement (EuroSPI 2014)*, 2014, vol. 425, pp. 182–193.
- [86] E. Mendes, “A systematic review of Web engineering research,” in *4th International Symposium on Empirical Software Engineering (ISESE' 05)*, 2005, pp. 498–507.
- [87] N. Salleh, E. Mendes, and J. Grundy, “Empirical studies of pair programming for {CS/SE} teaching in higher education: a systematic literature review,” *IEEE Trans. Softw. Eng.*, vol. 37, no. 4, pp. 509–525, 2011.
- [88] M. Brhel, H. Meth, A. Maedche, and K. Werder, “Exploring principles of user-centered agile software development: a literature review,” *Inf. Softw. Technol.*, vol. 61, no. C, pp.

163–181, 2015.

- [89] F. Selleri Silva, F. S. F. Soares, A. L. Peres, I. M. D. Azevedo, A. P. L. F. Vasconcelos, F. K. Kamei, and S. R. D. L. Meira, “Using {CMMI} together with agile software development: a systematic review,” *Inf. Softw. Technol.*, vol. 58, pp. 20–43, 2015.
- [90] J. E. Hannay, D. I. K. Sjøberg, and T. Dybå, “A systematic review of theory use in software engineering experiments,” *IEEE Trans. Softw. Eng.*, vol. 33, no. 2, pp. 87–107, 2007.
- [91] E. Hossain, M. A. Babar, and H. Paik, “Using {SCRUM} in global software development: a systematic literature review,” in *4th International Conference on Global Software Engineering (ICGSE 2009)*, 2009, pp. 175–184.
- [92] M. Unterkalmsteiner, T. Gorschek, C. K. Cheng, R. B. Permadi, and R. Feldt, “Evaluation and measurement of software process improvement -- a systematic literature review,” *IEEE Trans. Softw. Eng.*, vol. 38, no. 2, pp. 398–424, 2012.
- [93] D. S. Cruzes and T. Dyb\ra, “Research synthesis in software engineering: a tertiary study,” *Inf. Softw. Technol.*, vol. 53, no. 5, pp. 440–455, 2011.
- [94] L. M. Pickard, B. A. Kitchenham, and P. W. Jones, “Combining empirical results in software engineering,” *Inf. Softw. Technol.*, vol. 40, no. 14, pp. 811–821, 1998.
- [95] D. Galin and M. Acrahami, “Are {CMM} program investments beneficial? {A}nalyzing past studies,” *IEEE Softw.*, vol. 23, no. 6, pp. 81–87, 2006.
- [96] C. B. Seaman, “Qualitative methods,” in *Guide to Advanced Empirical Software Engineering*, Springer, 2008, pp. 35–62.
- [97] J. Miller, “Applying meta-analytical procedures to software engineering experiments,” *J. Syst. Softw.*, vol. 54, pp. 29–39, 2000.
- [98] G. W. Noblit and R. D. Hare, *Meta-ethnography: synthesising qualitative studies*. SAGE Publications, 1988.
- [99] M. Rodgers, A. Sowden, M. Petticrew, L. Arai, H. Roberts, N. Britten, and P. Jennie, “Testing methodological guidance on the conduct of narrative synthesis in systematic reviews,” *Evaluation*, vol. 15, no. 1, pp. 47–71, 2009.
- [100] W. Afzal, R. Torkar, and R. Feldt, “A systematic review of search-based testing for non-functional system properties,” *Inf. Softw. Technol.*, vol. 51, no. 6, pp. 957–976, 2009.

- [101] C. C. Ragin, *The comparative method: moving beyond qualitative and quantitative strategies*. University of California Press, 1987.
- [102] M. Dixon-Woods, S. Agarwal, D. Jones, B. Young, and A. Sutton, "Synthesising qualitative and quantitative evidence: a review of possible methods," *J. Heal. Serv. Res. Policy*, vol. 10, no. 1, pp. 45–53, 2005.
- [103] A. Davis, O. Dieste, A. Hickey, N. Juristo, and A. M. Moreno, "Effectiveness of requirements elicitation techniques: empirical results derived from a systematic review," in *14th IEEE International Conference on Requirements Engineering (RE 2006)*, 2006, pp. 179–188.
- [104] A. Nguyen-Duc, D. S. Cruzes, and R. Conradi, "The impact of global dispersion on coordination, team performance and software quality -- A systematic literature review," *Inf. Softw. Technol.*, vol. 57, pp. 277–294, 2015.
- [105] B. Glaser and A. Strauss, *The discovery of grounded theory: strategies for qualitative research*. Aldine Transaction, 1999.
- [106] J. Corbin and A. Strauss, *Basics of qualitative research*. SAGE Publications, 2014.
- [107] M. B. Miles, A. M. Huberman, and J. Salda\Hna, *Qualitative data analysis: an expanded sourcebook*. SAGE Publications, 2013.
- [108] R. Pawson, T. Greenhalgh, G. Harvey, and K. Walshe, "Realist review -- a new method of systematic review designed for complex policy interventions," *J. Health Serv. Res. Policy*, vol. 10, no. 1, pp. 21–34, 2005.
- [109] E. Barnett-Page and J. Thomas, "Methods for the synthesis of qualitative research: a critical review," *BMC Med. Res. Methodol.*, vol. 9, no. 59, pp. 1–11, 2009.
- [110] H. J.P.T. and G. S. (editors), *Cochrane handbook for systematic reviews of interventions - - Version 5.1.0 [updated March 2011]*. 2011.
- [111] N. B. Ali, K. Petersen, and C. Wohlin, "A systematic literature review on the industrial use of software process simulation," *J. Syst. Softw.*, vol. 97, no. C, pp. 65–85, 2014.
- [112] D. Weyns and T. Ahmad, "Claims and evidence for architecture-based self-adaptation: a systematic literature review," in *7th European Conference on Software Architecture (ECSA 2013)*, 2013, pp. 249–265.
- [113] E. Y. Nakagawa, M. Gonçalves, M. Guessi, L. B. R. Oliveira, and F. Oquendo, "The state

of the art and future perspectives in systems of systems software architectures," in *1st International Workshop on Software Engineering for Systems-of-Systems (SESoS 2013)*, 2013, pp. 13–20.

- [114] D. Moher and A. Tsertsvadze, "Systematic reviews: when is an update an update?," *Lancet*, vol. 367, no. 9514, pp. 881–883, 2006.
- [115] L. B. R. Oliveira, "Projeto arquitetural de sistemas robóticos orientados a serviços," Instituto de Ciências Matemáticas e de Computação - ICMC, Universidade de São Paulo - USP, 2015.
- [116] O. Dieste, M. Lopez, and F. Ramos, "Formalizing a systematic review updating process," in *8th International Conference on Software Engineering Research, Management and Applications (SERA 2008)*, 2008, pp. 143–150.
- [117] K. R. Felizardo, E. Y. Nakagawa, S. G. MacDonell, and J. C. Maldonado, "A visual analysis approach to update systematic reviews," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*, 2014, pp. 1–10.
- [118] S. J. Taylor, R. Bogdan, and M. DeVault, *Introduction to qualitative research methods: a guidebook and resource*. Wiley, 2015.
- [119] E. F. Souza, R. A. Falbo, and N. L. Vijaykumar, "Knowledge management initiatives in software testing: A mapping study," *Inf. Softw. Technol.*, vol. 57, no. 1, pp. 378–391, 2015.
- [120] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," in *12nd International Conference on Evaluation and Assessment in Software Engineering (EASE' 08)*, 2008, pp. 1–10.
- [121] B. A. Kitchenham, D. Budgen, and P. O. Brereton, "Using mapping studies as the basis for further research: a participant-observer case study," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 638–651, 2011.
- [122] D. Budgen, B. A. Kitchenham, S. C. Charters, M. Turner, O. P. Brereton, and S. G. Linkman, "Presenting Software Engineering Results using Structured Abstracts: A Randomised Experiment," *Empir. Softw. Eng.*, vol. 13, no. 4, pp. 435–468, 2008.
- [123] B. A. Kitchenham, P. O. Brereton, and D. Budgen, "Mapping study completeness and reliability -- a case study," in *16th International Conference on Evaluation Assessment in*

Software Engineering (EASE 2012), 2012, pp. 126–135.

- [124] S. Easterbrook, “Empirical research methods for software engineering,” in *22nd IEEE/ACM International Conference on Automated Software Engineering (ASE 2007)*, 2007, p. 574.
- [125] I. Nonaka and H. Takeuchi, *The knowledge-creating company. How japanese companies create the dynamics of innovation*. Oxford University Press, 1997.
- [126] J. Portillo-Rodríguez, A. Vizcaíno, M. Piattini, and S. Beecham, “Tools used in global software engineering: a systematic mapping review,” *Inf. Softw. Technol.*, vol. 54, no. 7, pp. 663–685, 2012.
- [127] O. Pedreira, F. Garcia, N. Brisaboa, and M. Piattini, “Gamification in software engineering -- a systematic mapping,” *Inf. Softw. Technol.*, vol. 57, no. 1, pp. 157–168, 2015.
- [128] A. D. Neto, R. Subramanyan, M. Viera, G. H. Travassos, and F. Shull, “Improving evidence about software technologies: a look at model-based testing,” *IEEE Softw.*, vol. 25, no. 6, pp. 242–249, 2008.