

Mapeamento Sistemático

Ricardo de Almeida Falbo

Estudos secundários visam revisar estudos primários relativos a certas questões de pesquisa, com o objetivo específico de integrar e sintetizar evidências relacionadas a essas questões. De maneira geral, o objetivo de um estudo secundário é prover a pesquisadores uma visão geral de uma área de pesquisa e ajudar a identificar lacunas na pesquisa nessa área (WHOLIN et al., 2013). Revisão Sistemática (RS) e Mapeamento Sistemático (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 particular conjunto de questões de pesquisa, tópico ou fenômeno de interesse, de uma maneira não tendenciosa e, até certo grau, repetível (KITCHENHAM; CHARTERS, 2007). O objetivo deste texto é apresentar conceitos básicos de MS. Para ilustrar alguns dos aspectos aqui discutidos, são citados alguns MSs e mostrado como eles lidaram com esses aspectos.

Este texto está estruturado da seguinte forma: a Seção 1 discute o que é um MS, quais as semelhanças e as diferenças entre um MS e uma RS, bem como razões para e problemas enfrentados ao se realizar um MS; a Seção 2 apresenta o processo de MS; a Seção 3 aborda como definir objetivo e as questões de pesquisa para um MS; a Seção 4 discute estratégias de busca, a definição de critérios para selecionar estudos relevantes em um MS e como decidir pela inclusão ou não de um artigo; a Seção 5 trata da extração e categorização de dados em um MS; a Seção 6 discute as ameaças à validade de um MS; e finalmente, a Seção 7 trata de como sintetizar e apresentar os resultados de um MS.

1. O que é um Mapeamento Sistemático?

Um Mapeamento Sistemático (MS) é uma revisão ampla dos estudos primários existentes em um tópico de pesquisa específico que visa identificar a evidência disponível nesse tópico. Assim, um MS é um estudo secundário¹ que tem como objetivo identificar e classificar a pesquisa relacionada a um tópico amplo de pesquisa (KITCHENHAM; CHARTERS, 2007).

Resultados de um MS ajudam a identificar lacunas nesta área, capazes de sugerir pesquisas futuras e prover um guia para posicionar adequadamente novas atividades de pesquisa (KITCHENHAM; CHARTERS, 2007; KITCHENHAM et al., 2011; PETERSEN et al., 2008). Assim, MSs visam prover uma visão geral de um tópico e identificar se há subtópicos nos quais mais estudos primários são necessários.

RSs e MSs, sendo estudos secundários, apoiam a identificação e agregação das evidências disponíveis para tratar certas questões de pesquisa e auxiliam na identificação dessas lacunas (KITCHENHAM et al., 2010). Neste sentido, RSs e MSs apoiam a tomada de decisão relacionada a uma pesquisa a ser desenvolvida e há muitas semelhanças entre eles. Contudo, há também diferenças, em especial no que se referem a objetivos, questões

¹ Vale observar que RSs e MSs podem ser usados também para realizar estudos terciários. Estudos terciários são revisões de estudos secundários relacionados a uma mesma questão de pesquisa. Uma revisão terciária é uma revisão sistemática de revisões sistemáticas para responder a uma questão mais ampla. Um bom exemplo de MS como estudo terciário é o trabalho de Kitchenham et al. (2010).

de pesquisa, estratégias de busca, processo de seleção, avaliação da qualidade, análise de dados e resultados obtidos (PETERSEN et al., 2015).

No que se refere aos objetivos, RSs tentam agregar os estudos primários em termos de seus resultados e investigar se esses resultados são consistentes ou contraditórios. RSs visam, portanto, sintetizar evidências, considerando, inclusive, a sua força (KITCHENHAM et al., 2011; PETERSEN et al., 2015). MSs, por sua vez, visam prover uma visão geral de um tópico de pesquisa mais amplo e identificar tanto clusters de estudos que podem ser apropriados para estudo mais detalhado a ser feito por meio de uma RS, quanto subtópicos em que mais estudos primários são necessários (KITCHENHAM et al., 2012). MSs estão focados na estruturação de uma área de pesquisa (PETERSEN et al., 2015) e visam apenas classificar a literatura relevante e agregar estudos em relação a categorias definidas (KITCHENHAM et al., 2011).

Kitchenham et al. (2010) consideram que RSs e MSs são tipos de revisões sistemáticas da literatura. Segundo eles, RSs convencionais agregam resultados relacionados a uma questão de pesquisa específica, enquanto MSs visam encontrar e classificar os estudos primários em um tópico. Assim, MSs têm questões de pesquisa de caráter mais geral e podem ser usados para identificar a literatura disponível antes de se realizar RSs convencionais. Os mesmos métodos de busca e extração de dados são usados, mas MSs apoiam-se mais na tabulação dos estudos primários em categorias definidas. Assim, um MS visa construir uma estrutura e esquema de classificação em um campo de interesse. A análise dos resultados enfoca frequências de publicações por categorias dentro desse esquema, ainda que diferentes facetas possam ser combinadas para responder questões de pesquisa mais específicas (PETERSEN et al., 2008).

De fato, as principais diferenças entre MS e RS residem no escopo e nos procedimentos de análise. O escopo de um MS é geralmente mais amplo e a análise e síntese mais superficiais do que em uma RS (WHOLIN et al., 2013). Para tal, a estratégia de busca deve ser menos restritiva, de modo permitir recuperar mais estudos (KITCHENHAM et al., 2011). Assim, de maneira geral, MSs envolvem mais trabalhos, a serem analisados mais superficialmente, enquanto RSs envolvem menos trabalhos, mas que devem ser analisados com maior profundidade.

Uma importante diferença de caráter prático é a forma como os estudos são avaliados. Enquanto em uma RS os estudos devem ser estudados em detalhes para que as evidências possam ser analisadas e contrastadas, em um MS a leitura dos estudos é mais superficial, visando classificar cada estudo nas categorias que estão sendo levadas em consideração.

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 doutorado têm em MSs um meio valioso e eficiente de iniciar suas atividades de pesquisa (BUDGEN et al. 2008; KITCHENHAM et al. 2010).

Para pesquisadores em geral, incluindo também estudantes, um MS pode prover um corpo de conhecimento e um ponto de partida para suas pesquisas (inclusive de outros pesquisadores que não aqueles que realizaram o MS), ao invés de forçá-los a começar do nada (BUDGEN et al. 2008).

Algumas razões típicas para se realizar MSs incluem:

- Para examinar a extensão e a natureza de uma atividade de pesquisa (ARKSEY e O'MALLEY, 2005).

- Avaliar o valor de se realizar uma RS completa, determinando o potencial esforço necessário para tal (ARKSEY e O'MALLEY, 2005).
- Para coletar e sumarizar a pesquisa existente em um tópico. 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 tema de seu trabalho.
- 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. Primeiro, um MS pode ser conduzido, 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 (KITCHENHAM et al., 2011). Em uma abordagem na qual um MS precede uma RS, é possível apontar vários benefícios, dentre eles (KITCHENHAM et al., 2011):

- Reduz o tempo necessário para atividades de pesquisa subseqüentes, em especial na realização de uma RS.
- Fica mais fácil compreender a literatura e definir questões de pesquisa.
- O MS provê um conjunto de estudos que pode ser usado como grupo de controle para avaliar a string de busca da RS.
- Procedimentos, formulários e experiência podem ser reutilizados.

Além disso, um MS pode ser usado (KITCHENHAM et al., 2011):

- Como uma *baseline* contra a qual tendências de pesquisa podem ser rastreadas ao longo do tempo.
- Para justificar a realização de mais estudos primários quando houver poucos estudos empíricos relevantes (ou nenhum).
- Como um meio de identificar trabalhos relacionados ao que está sendo desenvolvido. Os resultados desses estudos podem ser usados para comparação com os resultados obtidos no trabalho sendo desenvolvido.
- Como um recurso educacional para se aprender sobre o tópico abordado.

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. Ele deve ter sido baseado em um processo de busca e seleção rigoroso, incluindo busca automática a fontes críticas, *snowballing* de referências encontradas nos estudos selecionados e busca ou comunicação direta com importantes pesquisadores e grupos de pesquisa, bem como em um sistema de classificação confiável e bem definido.

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. Mesmo para um trabalho de doutorado, este tempo pode ir além do estimado para a revisão da literatura. Em compensação, um MS de qualidade é uma excelente oportunidade de publicação. Para um trabalho de mestrado, é importante avaliar com cuidado o tópico escolhido, pois pode ser inviável realizar um MS completo. Restringir a busca a certas bases pode ser uma alternativa.

Dois problemas recorrentes em MSs são a seleção e a classificação. Ambos são um desafio, sobretudo para estudantes ou pesquisadores pouco familiarizados com o tema. Eles podem não estar familiarizados totalmente com o tópico em estudo e pode lhes faltar um entendimento mais sólido da terminologia empírica empregada nos artigos. Mesmo pesquisadores experientes podem fazer avaliações e classificações substancialmente diferentes de um mesmo trabalho. Normalmente, este 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. Entretanto, isso pode levar a carga significativa de trabalho para o orientador, tornando inviável o trabalho (BUDGEN et al. 2008).

Um desafio ainda maior é avaliar a qualidade dos estudos selecionados. Ainda que esta não seja uma avaliação obrigatória para MSs, quando realizada, impõe muitas dificuldades, começando pela definição dos critérios (objetivos) de qualidade e passando para as avaliações dos estudos primários. Esse desafio é ainda maior devido ao fato de muitos estudos, por serem apresentados em artigos de tamanho relativamente pequeno, apresentarem poucas informações sobre o trabalho realizado, em particular no que concerne à parte empírica (BUDGEN et al. 2008).

2. O Processo de Revisão/Mapeamento Sistemático

O processo de RS/MS envolve três fases principais: Planejamento da Revisão, Condução da Revisão e Publicação dos Resultados (KITCHENHAM; CHARTERS, 2007). Essas fases, bem como suas atividades, são conduzidas de modo iterativo. A Figura 1 mostra as fases e atividades do processo de RS/MS, as quais são detalhadas a seguir.

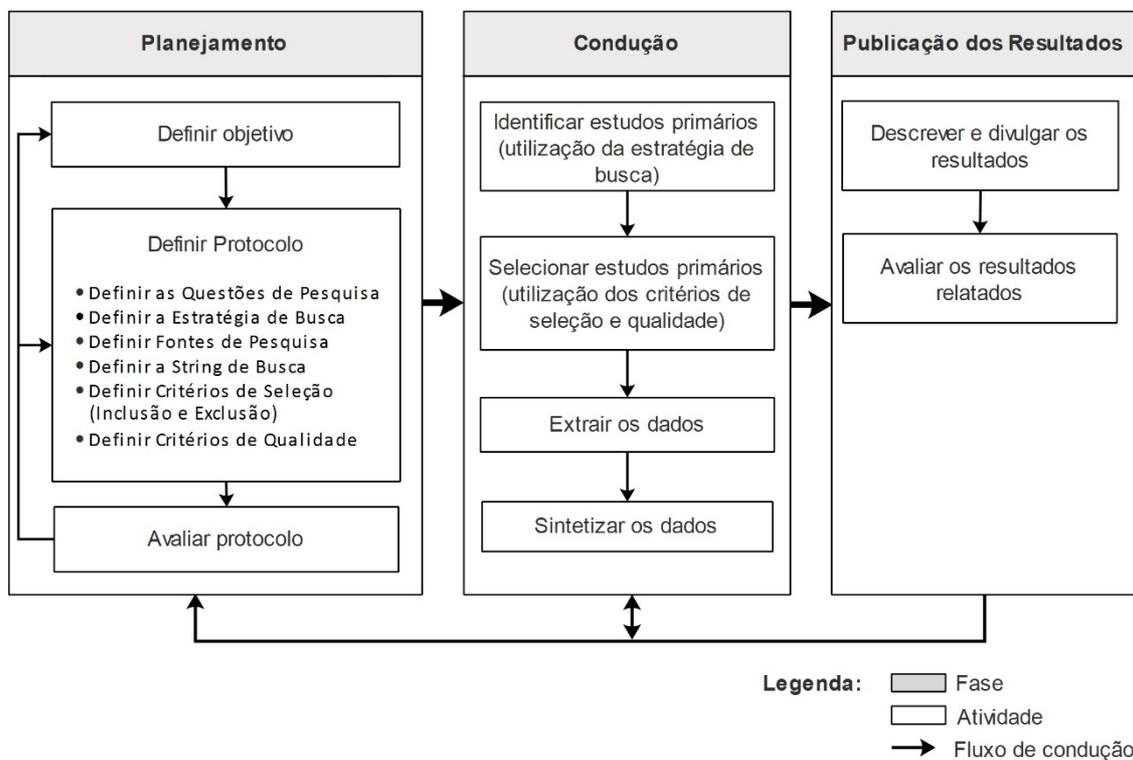


Figura 1 - Fases e Atividades do Processo da RS/MS.

Planejamento da Revisão

A fase de planejamento tem como objetivo identificar a real necessidade, ou seja, a motivação para a execução de um MS/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, outros MSs/RSs (KITCHENHAM; CHARTERS, 2007). Revisões terciárias podem ser executadas utilizando o mesmo processo adotado para uma RS. 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 de um MS/RS.

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 especifica as questões de pesquisa, a estratégia que será utilizada para conduzir o MS/RS, os critérios para a seleção dos estudos, e como os dados serão extraídos dos estudos e sintetizados.

Vale destacar que a qualidade do protocolo impacta diretamente a qualidade do MS/RS. Dessa forma, o protocolo deve ser validado antes que se prossiga com o MS/RS (KITCHENHAM; CHARTERS, 2007). Essa validação 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 identificar modificações que sejam necessárias. Para o teste piloto é preciso definir 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 do MS/RS, por meio da sugestão de especialistas da área ou combinando ambas as estratégias.

Em função da natureza iterativa do processo, os itens que compõem o protocolo podem ser refinados. P.ex., como é possível observar na Figura 1, durante a seleção dos estudos, novos termos não inicialmente considerados na *string* de busca podem ser identificados, levando ao refinamento da mesma e, posteriormente, a uma nova busca.

Condução da Revisão

Após a validação do protocolo, inicia-se a fase de condução do MS/RS. Durante essa fase, os estudos primários são identificados utilizando-se a estratégia de busca definida no protocolo. Uma vez identificados, os estudos precisam ser selecionados por meio da aplicação de critérios de seleção (critérios de inclusão e de exclusão), bem como os critérios de qualidade (mais aplicáveis a RSs).

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 a adoção de métodos que conduzam aos objetivos propostos na RS. 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 (KITCHENHAM; CHARTERS, 2007).

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 as questões de pesquisa da revisão e para facilitar posteriormente as análises e síntese dos resultados.

Publicação dos Resultados

A última fase do processo de MS/RS é a escrita dos resultados da RS, que poderiam ser divulgados aos potenciais interessados (KITCHENHAM; CHARTERS, 2007). Esses resultados podem ser divulgados por meio de relatórios técnicos, artigos de periódicos ou conferências, em capítulos de livros ou como uma seção de um trabalho de conclusão de curso (p.ex., uma tese de doutorado ou uma dissertação de mestrado). 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.

Vale destacar que todas as informações produzidas nas etapas de planejamento e condução devem ser adequadamente registradas para permitir a posterior publicação dos resultados do mapeamento. Para tal, é imprescindível usar ferramentas como planilhas eletrônicas e ferramentas de software para o gerenciamento de referências para registrar tais informações. As figuras 2 e 3 mostram um modelo de planilha eletrônica que pode ser usado para registrar as informações do planejamento e da seleção inicial dos estudos. A segunda planilha, com pequenas adaptações, pode ser usada também para o registro da seleção final. Uma opção ainda mais interessante é usar uma ferramenta de software de apoio à realização de MSs/RSs, tal como a ferramenta Start (disponível em http://lapes.dc.ufscar.br/tools/start_tool).

Mapeamento Sistemático - Informações Gerais		
Objetivo:		
Questões de Pesquisa		Rationale
RQ1	Quando e onde os estudos foram publicados?	
RQ2	Que tipos de pesquisa tem sido feitas?	
RQ3		
RQ4		
RQ5		
RQ6		
String de Busca:		
Período:		
Área:		
Tipo de Publicação:		
Fontes:	IEEE Xplore ACM Digital Library Scopus Science Direct Compendex ISI of Knowledge	
Critérios de Inclusão		
CI1		
Critérios de Exclusão		
CE1	O artigo não tem um resumo (abstract)	
CE2	A publicação é apenas um resumo (abstract)	
CE3	O artigo não está escrito em inglês	
CE4	O artigo é uma cópia ou uma versão mais antiga de um outro artigo já considerado	
CE5	O artigo não é um estudo primário (tais como editoriais, summaries of keynotes, tutorials)	
CE6	Não foi possível ter acesso ao trabalho completo.	
Artigos de Controle		Fontes
AC1	Dados do artigo	
AC2		
AC3		

Figura 2 - Modelo de Planilha com informações relativas ao Planejamento de um MS

Mapeamento Sistemático - Resultado da 1ª Etapa (Eliminação de Duplicatas) e Aplicação do 1º Filtro (2ª Etapa)

Nesta planilha constam os artigos a serem avaliados pela aplicação do primeiro filtro. Foram removidos os artigos repetidos e adicionado um ID para cada artigo.

Cada artigo devidamente identificado deve ser analisado, levando-se em conta os critérios de seleção (CI e CEs) aplicados considerando apenas título, abstract e palavras-chave. Caso a análise aponte na coluna "Análise" deve ser indicado o critério de exclusão aplicado ("CI ou CE1 a CE5), conforme critérios apontados na planilha "Geral".

ID	Bases	Ano	Título	Autores	Abstract
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Figura 3 - Modelo de Planilha para a Seleção Inicial de Estudos.

As seções que se seguem detalham algumas das atividades do processo de MS/RS, tendo como foco MSs e apresentando um exemplo de MS sobre o seguinte tema: Gestão do Conhecimento em Teste de Software, publicado em (SOUZA et al., 2015).

3. Objetivo e Questões de Pesquisa

A primeira coisa a se fazer é definir o objetivo do MS. De maneira geral, o propósito de um MS é revisar um tópico amplo de pesquisa e classificar os estudos primários relacionados a esse tópico.

Definido o objetivo do mapeamento, é possível iniciar a definição das questões de pesquisa. De maneira geral, as questões de pesquisa para um MS são questões de nível mais alto (e, portanto, menor profundidade). Alguns aspectos tipicamente discutidos em um MS levam a questões como: Quais subtópicos têm sido tratados? Que métodos de pesquisa têm sido usados? Em que veículos de publicação esses estudos têm sido reportados? Além disso, MSs tipicamente têm um número maior de questões de pesquisa, se comparados com RSs. Em suma, as questões de pesquisa em um MS são mais gerais e em maior quantidade, na medida em que elas buscam descobrir tendências de pesquisa (p.ex., tendência de publicação ao longo do tempo, tópicos cobertos na literatura etc.).

Petersen et al. (2015) apontam cinco aspectos presentes em questões de pesquisa de muitos 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. (2006) e revisada por Petersen et al. (2015), a qual 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 (p.ex., academia, indústria, governo, projeto ou organização).
- Veículo de publicação: refere-se ao tipo do veículo de publicação onde os estudos têm sido publicados (periódicos, conferências e workshops), bem como aos

veículos específicos (os periódicos, as conferências e os workshops específicos) que têm sido alvo das publicações.

- Método de pesquisa: refere-se aos métodos científicos usados (p.ex., estudo de caso, experimento, *survey* etc.).

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, combinando os resultados dessa questão relativa ao tempo com os resultados de uma questão sobre o tipo de pesquisa, por exemplo, é possível avaliar se a maturidade da pesquisa no tópico tem aumentado ao longo dos anos.

A seguir são apresentados o objetivo e as questões de pesquisa definidos no contexto do mapeamento apresentado em (SOUZA et al., 2015).

Gestão do Conhecimento em Teste de Software
Objetivo: Prover uma visão geral do estado corrente da pesquisa relacionada a iniciativas de gestão de conhecimento aplicada ao domínio de teste de software, investigando problemas tratados por tais iniciativas, benefícios e fraquezas, escopo alvo das iniciativas e tecnologias empregadas.
Questões de Pesquisa: <ul style="list-style-type: none">• QP1: Quando e onde os estudos têm sido publicados?• QP2: Do ponto de vista de teste de software, que aspectos têm sido enfocados?• QP3: Do ponto de vista de gestão do conhecimento, que tópicos têm sido enfocados?• QP4: Que tipos de pesquisa têm sido feitas?• QP5: Que 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?

4. Estratégias de Busca e Seleção

A estratégia de busca a ser usada é um ponto chave para o sucesso ou fracasso de um MS. Ela inclui a definição de: (i) métodos de busca, i.e., como as buscas serão realizadas; (ii) fontes de pesquisa, ou seja, os locais onde os estudos serão procurados; (iii) *string* de busca a ser usada; (iv) critérios de seleção; e (v) etapas do processo de seleção a serem realizadas.

Método de Busca

No que se refere ao método de busca em si, há três métodos principais que podem ser aplicados: busca automática em bases de dados digitais, busca manual e bola de neve (*snowballing*). De maneira geral, para que um mapeamento proveja uma visão ampla do tópico, é necessário realizar buscas automáticas. Assim, em MSs, busca manual e *snowballing* são consideradas estratégias complementares a serem usadas em conjunto com a busca automática.

A busca manual deve ser usada quando se sabe que algum evento ou periódico específico é um importante veículo de publicação para o tópico em estudo e o mesmo não está contemplado nas bases selecionadas para a busca automática.

Snowballing é muito útil para encontrar estudos relevantes a partir dos estudos selecionados por meio da aplicação de outros métodos ou por meio de *snowballing* realizado previamente. Contudo, só deve ser aplicado em casos em que poucos estudos são selecionados com outros métodos, caso contrário pode ser inviável do ponto de vista prático.

Uma quarta abordagem que pode ser aplicada consiste em, uma vez selecionados os estudos aplicando-se os métodos anteriormente discutidos, identificar autores e grupos de pesquisa que têm atuado na área e fazer uma busca por outros trabalhos desses autores/grupos de pesquisa em suas páginas institucionais. Para tornar esta busca mais sistemática, recomenda-se utilizar as próprias bibliotecas digitais usadas como fonte, DBLP² (*Digital Bibliography & Library Project*), um repositório bibliográfico contendo as principais publicações dos pesquisadores em Ciência da Computação, o Google Acadêmico³, ou redes sociais de pesquisa como Research Gate ou Academia. No caso de se utilizar o Google Acadêmico, deve-se desmarcar a opção “incluir citações”, caso contrário, além dos trabalhos dos autores, serão mostrados aqueles trabalhos que citam os trabalhos dos autores.

Vale destacar que aplicar múltiplas abordagens pode ser útil por um lado, mas, por outro, pode consumir muito tempo. Assim, nem sempre é prático aplicar vários métodos e se recomenda selecioná-los 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.

Fontes de Pesquisa

Em relação às fontes de pesquisa, no que concerne à busca automática, é importante distinguir entre bibliotecas digitais de editoras (tais como IEEE Xplorer, ACM Digital Library⁴ e Springer Link) e indexadores (p.ex., Scopus e Web of Science). As bibliotecas digitais de editoras contêm basicamente os artigos publicados apenas por elas, enquanto os indexadores indexam artigos de várias bibliotecas digitais. Poder-se-ia pensar que basta, então, fazer a pesquisa em indexadores abrangendo as principais bibliotecas digitais. Contudo, nem sempre os indexadores cobrem completamente o conteúdo das bibliotecas digitais e, muitas vezes, alguns artigos só são retornados a partir das bibliotecas digitais das próprias editoras. Este tem sido o caso, sobretudo, da IEEE

² <http://dblp.uni-trier.de/>

³ <https://scholar.google.com.br/>

⁴ A ACM Digital Library é, na verdade, um mecanismo híbrido, na medida em que, além dos artigos publicados pela ACM, ela também indexa trabalhos de outras editoras, mas de forma limitada.

Xplorer e ACM Digital Library. Neste sentido, recomenda-se usar como fonte de pesquisa para a busca automática tanto bibliotecas digitais de editoras quanto indexadores.

A escolha de quais fontes de pesquisa utilizar em um MS está diretamente relacionada ao esforço a ser empreendido no mapeamento. Quanto maior a quantidade de fontes de pesquisa, maior é a chance de se obter uma cobertura abrangente dos trabalhos publicados. Contudo, o trabalho a ser realizado também crescerá bastante. Assim, é necessário avaliar a relação custo-benefício de se usar muitas bases de dados.

Segundo Dyba et al. (2007), o uso das bibliotecas digitais da IEEE (IEEE Xplorer) e da ACM (ACM Digital Library), combinado com duas bases de dados de indexadores (p.ex., Scopus e Engineering Village) pode ser suficiente. Ainda que esta diretriz possa ser útil, recomenda-se avaliar se esta combinação é efetivamente suficiente, tendo em vista que o comportamento dos mecanismos de busca é alterado por atualizações. Assim, é importante avaliar a cobertura obtida e recomenda-se avaliar a inclusão das seguintes fontes de dados: IEEE Xplorer, ACM Digital Library, Springer Link, Scopus, Web of Science, Engineering Village e Science Direct. Em todos os MSs por nós realizados nos últimos tempos que usaram todas essas fontes de dados, houve casos de artigos que retornaram somente em uma dessas bases (para todas as bases), o que demonstra que a diretriz de Dyba et al. não garante a mesma cobertura que a obtida usando todas as bases.

De maneira geral, recomendamos que MSs feitos no contexto de teses de doutorado, que requerem uma cobertura mais abrangente, sejam feitos considerando todas as bases listadas acima. Já para MSs feitos no contexto de dissertações de mestrado, a diretriz provida por Dyba et al. (2007) pode ser aplicada.

Deve-se evitar o uso de mecanismos de busca gerais (como o Google), pois eles não apresentam o mesmo comportamento ao longo tempo, comprometendo, assim, a possibilidade de se reproduzir as buscas realizadas. O mesmo ocorre com o Google Acadêmico (Google Scholar) e, portanto, esta base de dados não deve ser considerada como fonte de pesquisa para um MS. Isso não impede que estes mecanismos sejam usados para encontrar estudos retornados na busca. Muito pelo contrário. Uma prática bastante comum é fazer a busca automática nas bibliotecas digitais das editoras e nos indexadores e, uma vez selecionados os estudos, procurar seus textos completos usando os mecanismos mais gerais.

Vale destacar que cada base de dados possui características e limitações próprias e que as máquinas de busca trabalham de maneiras diferentes. Assim, a *string* de busca deve ser adaptada para rodar adequadamente em cada base (WHOLIN et al. 2013). 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 (p.ex., título, resumo e palavras-chave) ou se as buscas são feitas sempre considerando o texto completo etc.

De fato, é importante ressaltar que a maior parte das bases anteriormente mencionadas não são projetadas para apoiar eficientemente a realização de MSs/RSs, apresentando problemas que dificultam a realização de buscas automáticas, tais como:

- Algumas bibliotecas digitais não permitem limitar a busca considerando somente título, resumo e palavras-chave. Este é o caso da SpringerLink. Essa limitação reduz a precisão da busca e faz com que uma grande quantidade de estudos irrelevantes seja retornada, dificultando ou até mesmo inviabilizando o seu uso como fonte para um MS. Quando tal base é usada, é necessário executar programas específicos (scripts), desenvolvidos para resolver esse problema, o que nem sempre é fácil e confiável.

- Algumas bases, como a Scopus, usam como palavras-chave, além daquelas fornecidas pelos autores, outras extraídas de processamento dos textos, ditas palavras de indexação (*indexing words*), as quais são consideradas quando se faz busca por palavra-chave. O uso de tais palavras-chave pode, muitas vezes, levar novamente a uma grande quantidade de estudos irrelevantes, dificultando ou até mesmo inviabilizando a busca por palavra-chave, quando a *string* de busca envolve termos bem gerais como “*Software Engineering*” ou “*User Interface*”.
- Em alguns casos, como no da IEEE Xplore, o plural de cada termo deve ser informado. Por isso, antes da utilização de uma base como fonte, deve-se verificar a necessidade de especificar ou não termos no plural.
- Algumas bases impõem limitações quanto ao número de caracteres ou de termos. Este é o caso da IEEE Xplorer. 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.

Gestão do Conhecimento em Teste de Software
Estratégias de Busca Aplicadas: Busca Automática, <i>snowballing</i> e pesquisa direta a autores/grupos de pesquisa (usando as bibliotecas digitais consultadas anteriormente, DBLP e páginas pessoais dos pesquisadores).
Fontes: <ul style="list-style-type: none"> • IEEE Xplore (http://ieeexplore.ieee.org). • ACM Digital Library (http://dl.acm.org). • SpringerLink (http://www.springerlink.com). • Scopus (http://www.scopus.com). • Science Direct (http://www.sciencedirect.com). • Compendex (http://www.engineeringvillage2.org). • ISI of Knowledge (http://www.isiknowledge.com).

***String* de Busca**

Em relação à definição da *string* de busca de um MS, várias abordagens podem ser usadas, tais como (PETERSEN et al. 2015): consultar especialistas, melhorar iterativamente a *string*, identificar termos a partir de artigos conhecidos e usar padrões, enciclopédias e tesouros. Vale destacar que essas abordagens não são difíceis de aplicar, não consomem muito tempo e melhoram bastante a qualidade da *string* de busca (PETERSEN et al., 2015). Assim, recomenda-se aplicá-las conjuntamente.

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 relativos a um mesmo aspecto, que podem ser considerados sinônimos, concatenando-os com o conectivo OU (OR em inglês). Posteriormente, cada grupo de termos é concatenado com os demais por meio de conectivos E (AND em inglês).

Vale ressaltar que os termos da *string* de busca devem estar alinhados ao objetivo e às questões de pesquisa do MS. Esta preocupação pode ser claramente percebida no exemplo discutido acima, no qual a *string* de busca é montada completamente alinhada com o objetivo do mapeamento e as questões de pesquisa.

Em qualquer caso, a *string* de busca precisa ser avaliada. Especialistas no tema podem ser chamados a avaliar a *string*, mas somente isso é insuficiente. Uma forma muito comum e eficiente de se avaliar a *string* de busca é utilizar um conjunto de estudos primários conhecidos, dito grupo de controle. Neste caso, a *string* de busca é rodada em uma das bases de dados previamente selecionadas (na qual se tenha certeza de que os artigos do grupo de controle estão presentes) e se verifica se os artigos de controle retornam. Este procedimento é chamado de teste piloto e serve, na verdade, como um meio de se avaliar todo o protocolo do MS. Para a formação do grupo de controle, recomenda-se a participação de especialistas no tópico do MS, acompanhada de uma revisão informal da literatura. Como base a ser usada para o teste piloto, recomenda-se a Scopus⁵, visto que ela é bastante abrangente, seu mecanismo de busca bastante eficiente e fácil de usar, bem como provê muitas facilidades úteis para se aperfeiçoar a *string*, tais como facilidades de diferenças entre buscas.

Na avaliação da *string* de busca deve-se considerar a precisão da mesma. Se a *string* de busca retornar um número muito grande de trabalhos irrelevantes, então vale a pena tentar refiná-la de modo a restringi-la um pouco mais. A meta é obter uma *string* que retorne um grande número de estudos relevantes, associado ao menor número possível de estudos irrelevantes. Ou seja, retornar mais artigos não é necessariamente melhor, mas sim retornar um volume menor de artigos contendo mais artigos relevantes. 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 em 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.

Gestão do Conhecimento em Teste de Software
<p>Foram considerados dois grupos de termos, um relativo a Gestão do Conhecimento e outro relativo a Teste de Software, conforme abaixo:</p> <p>Gestão de Conhecimento: “<i>knowledge management</i>” OR “<i>knowledge reuse</i>” OR “<i>knowledge sharing</i>” OR <i>knowledge transfer</i>”</p> <p>Teste de software: “<i>software testing</i>” OR “<i>software test</i>” OR (“<i>software project</i>” AND (“<i>test</i>” OR “<i>testing</i>”))</p>
<p><i>String</i> de Busca:</p> <p>(“<i>software testing</i>” OR “<i>software test</i>” OR (“<i>software project</i>” AND (“<i>test</i>” OR “<i>testing</i>”))) AND (“<i>knowledge management</i>” OR “<i>knowledge reuse</i>” OR “<i>knowledge sharing</i>” OR “<i>knowledge transfer</i>”)</p>

CrITÉRIOS de Seleção

A definição de critérios de seleção é fundamental para garantir a qualidade nos resultados obtidos em um MS, estabelecendo características que um estudo deve conter para ser considerado relevante no contexto do MS (critérios de inclusão) e características que levam à exclusão de estudos que não obedecem aos critérios definidos (critérios de exclusão).

⁵ <http://www.scopus.com>

Os critérios de inclusão definem características que um estudo deve conter para ser considerado relevante no contexto do MS. Por outro lado, os critérios de exclusão estabelecem características para excluir estudos que não obedecem ao contexto definido. Esses critérios são criados com base nas questões de pesquisa 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 do MS. 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). Além desse, pode ser definido um conjunto de critérios de exclusão gerais que vêm sendo adotados em MSs publicados 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 (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 reprodutibilidade 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. Neste caso, mesmo que um estudo tenha sido selecionado na seleção inicial (aquela que considera apenas título, resumo e palavras-chave), 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 do MS. 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:** Um MS é um estudo secundário. Ela 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, anais de eventos (congressos, workshops, simpósios, etc.), tutoriais e outros estudos secundários (outros MSs, revisões sistemáticas da literatura ou mesmo artigos apresentando revisões informais da literatura) e partes ou capítulos de livros. No caso de capítulos de livros, deve-se ter atenção especial. Atualmente, é bastante comum encontrar livros que, na verdade, são uma coletânea de trabalhos de pesquisa. Nesses casos, capítulos individuais devem sim ser selecionados, obviamente, quando relatarem um estudo primário; e
- **Não foi possível ter acesso ao estudo:** A maior parte das bibliotecas digitais 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 na Web. Mas o mesmo não ocorre com o texto completo. Por exemplo, pesquisadores brasileiros com acesso apenas ao Portal de Periódicos da CAPES⁶ não têm acesso a estudos publicados na Springer Link. 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 bom que o revisor use outros meios para tentar ter acesso ao estudo, tais como solicitar por e-mail aos autores o envio do estudo, buscar em redes sociais acadêmicas voltadas para pesquisadores, tais como Research Gate⁷ e Academia⁸, ou procurar encontrar o artigo na Web usando mecanismos de busca gerais como o Google.

Os critérios de seleção devem ser avaliados através de um teste piloto. 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 aos estudos avaliados antes da modificação.

Gestão do Conhecimento em Teste de Software

Foram utilizados os critérios de exclusão citados anteriormente e um único critério de inclusão: O estudo discute uma iniciativa de gestão de conhecimento em teste de software.

Processo de Seleção

A Figura 4 mostra as etapas típicas do processo de seleção de estudos para um MS. Um mesmo estudo pode estar indexado em mais de uma base, de modo que existirão estudos duplicados. Assim, antes de iniciar efetivamente o processo, a duplicidade de estudos deve ser eliminada. O processo de seleção efetivamente é geralmente realizado em duas etapas: seleção inicial e seleção final.

Na seleção inicial, os critérios de inclusão e exclusão são aplicados em todos os estudos identificados, por meio da avaliação de seus resumos. Muitas vezes é difícil

⁶ <http://www.periodicos.capes.gov.br>

⁷ <http://www.researchgate.net>

⁸ <http://www.academia.edu>

identificar se um estudo é ou não relevante apenas com a leitura do resumo. Assim, na dúvida sobre a inclusão ou não de um estudo durante a primeira etapa, deve-se optar pela sua inclusão, sendo a decisão pela permanência tomada durante a segunda etapa. Já durante a seleção final, os critérios de inclusão e exclusão são aplicados nos estudos incluídos na primeira etapa, por meio da avaliação de seus textos completos.

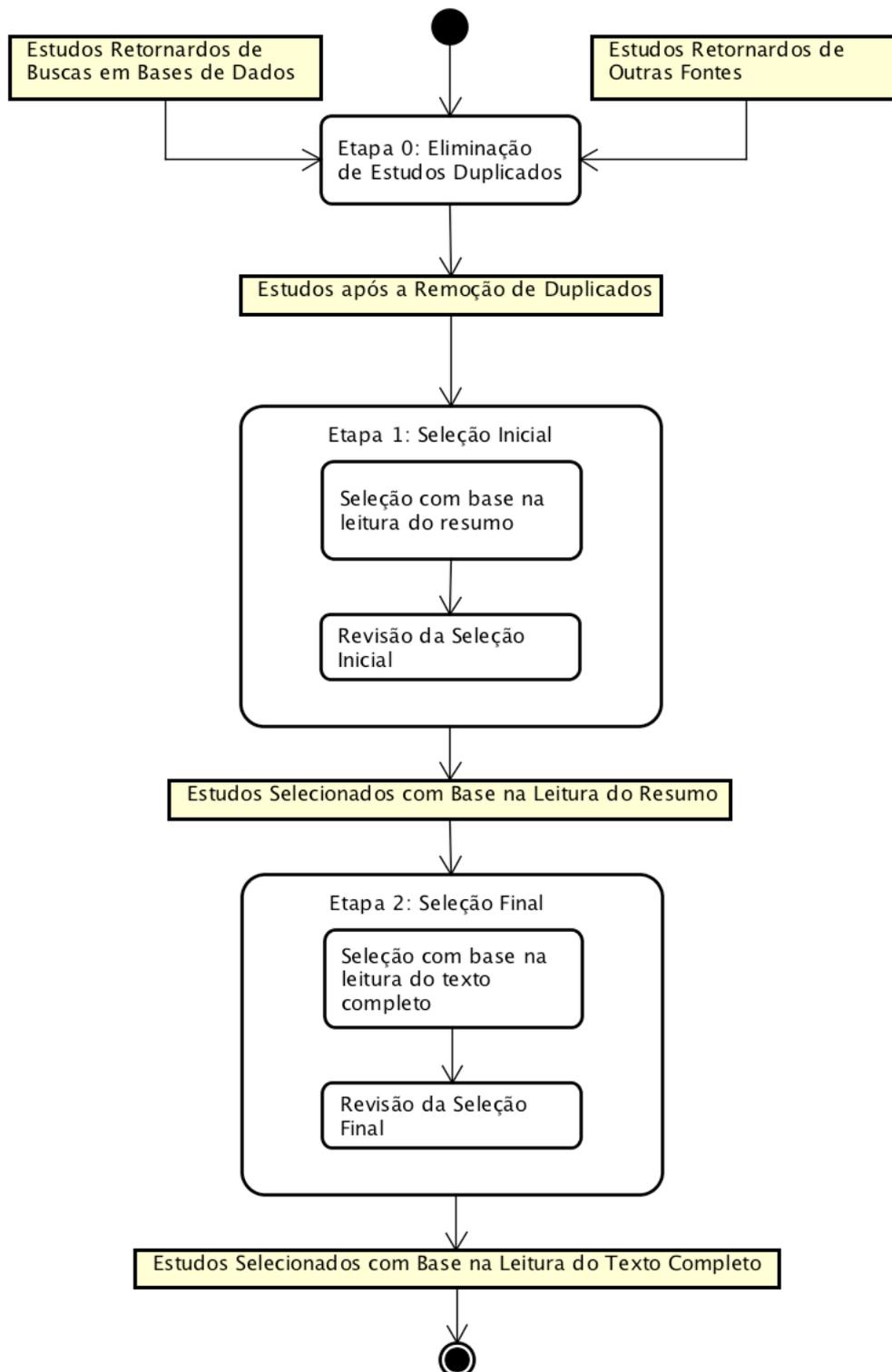


Figura 4 - Etapas do processo de seleção de estudos.

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 do processo de seleção e visa evitar que estudos relevantes sejam excluídos, uma vez que excluir estudos que devem ser avaliados e sintetizados pode comprometer o MS. Quando a atividade de seleção for realizada por dois ou mais pesquisadores de forma independente, estudos podem ser classificados de forma divergente. Durante a atividade de revisão da seleção, esses estudos devem ser reavaliados pelos próprios pesquisadores, ou até mesmo com a ajuda de um pesquisador externo. No caso de pesquisadores que conduzem a atividade de seleção de forma isolada ou quando a seleção é dividida entre os pesquisadores, mas cada trabalho é avaliado inicialmente apenas por um deles, a atividade de revisão da seleção serve para que os pesquisadores possam discutir suas decisões com outros pesquisadores. Pelo menos 1/3 da seleção de estudos deve ser revisada. É importante que os pesquisadores marquem explicitamente aqueles estudos que são objetos de dúvida, os quais devem ser obrigatoriamente revisados. O restante dos estudos (para se atingir pelo menos 1/3) deve ser escolhido aleatoriamente. Se a quantidade de estudos retornados não for muito grande e houver vários pesquisadores envolvidos com o processo de seleção, esse percentual pode aumentar, e até mesmo atingir a totalidade dos estudos retornados.

A seleção pode envolver diferentes configurações de participantes, que podem ser sintetizadas em quatro configurações principais como ilustra a Figura 5.

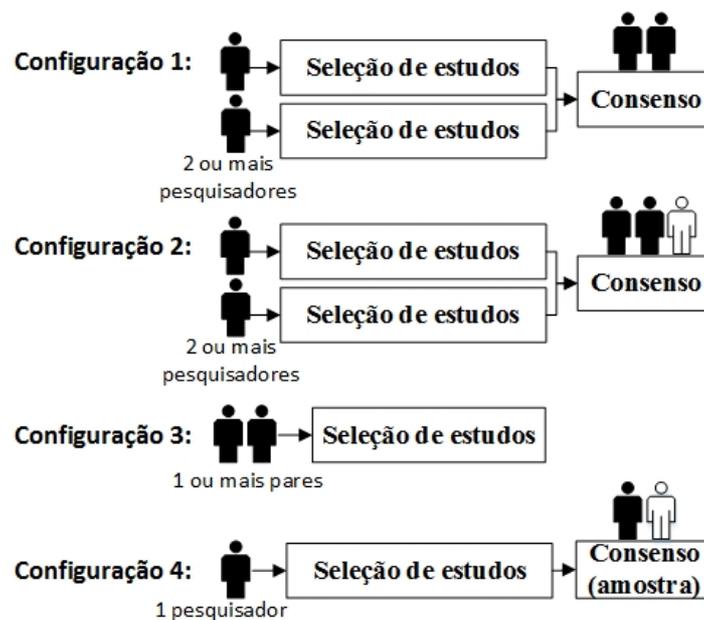


Figura 5 – Configurações típicas na seleção de estudos.

Na Configuração 1, dois ou mais pesquisadores avaliam todos os estudos de forma independente e comparam os resultados. Conflitos são solucionados por consenso entre os próprios pesquisadores. Na Configuração 2, todos os estudos são avaliados por dois (ou mais) pesquisadores de forma independente. Um pesquisador externo avalia as atividades conduzidas, classificando os estudos com divergências. Na Configuração 3, dois pesquisadores realizam a atividade de seleção em conjunto, ou seja, eles trabalham juntos na seleção de cada estudo, já resolvendo divergências e, portanto, não há uma etapa de consenso posterior. Finalmente, na Configuração 4, um pesquisador realiza a atividade de seleção e um segundo pesquisador (ou vários pesquisadores) avalia o resultado. Conflitos são solucionados por consenso. Esta configuração é a mais usada em casos de

dissertações de mestrado e teses de doutorado em que apenas o estudante realiza a seleção e seus orientadores (e outros pesquisadores) realizam a revisão.

Nas configurações que contam com a participação de um pesquisador externo, a condição básica para atuar como avaliador é ser um especialista no tema da revisão. Em casos de dissertações de mestrado e teses de doutorado, é muito comum que os orientadores desempenhem o papel de avaliador, mas outros pesquisadores com conhecimento sobre o tema do MS também podem atuar como pesquisadores externos.

Variações ou combinações das configurações ilustradas na Figura 5 são possíveis. Uma forma muito comum de trabalhar em equipe consiste em dividir os estudos a serem avaliados pelo conjunto de pesquisadores. Cada pesquisador faz a seleção de parte dos estudos, tendo seu trabalho revisado por outro pesquisador da equipe. Pesquisadores externos e os demais pesquisadores envolvidos no MS são chamados para resolver as divergências que persistam e avaliar aleatoriamente outras decisões, até se obter a cobertura mínima de 1/3 dos estudos.

Para resolver divergências sobre a inclusão de um estudo, é aconselhável obter informações adicionais sobre o trabalho. 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.

Decidir pelos estudos a serem incluídos em um MS envolve julgamento por parte dos pesquisadores envolvidos. Na seleção inicial (feita pela leitura dos resumos), cada revisor envolvido deve avaliar o estudo como "incluído", "excluído" ou "em dúvida". Conforme discutido anteriormente, pelo menos 1/3 dos estudos deve ter pelo menos duas avaliações. Estudos avaliados como duvidosos por um pesquisador devem ser obrigatoriamente revisados por outro. De posse dos julgamentos dos diferentes revisores, eles devem ser combinados. Isso pode ser feito de diferentes maneiras. Uma delas consiste em combinar as avaliações dos dois revisores usando regras de decisão (RDs), como as mostradas abaixo:

- RD1. Incluído + Incluído = Incluído
- RD2. Incluído + Em Dúvida = Incluído
- RD3. Incluído + Excluído = Incluído
- RD4. Em Dúvida + Em Dúvida = Incluído
- RD5. Em Dúvida + Excluído = Excluído
- RD6. Excluído + Excluído = Excluído

As regras de decisão RD4 e RD5 podem ser alteradas para tornar o conjunto de regras mais ou menos inclusivo. P.ex., a regra RD5 poderia ser alterada para "Em Dúvida + Excluído = Incluído". Neste caso, teríamos o conjunto de regras mais inclusivo sendo aplicado. Por outro lado, mantendo-se a proposta original de RD5 e alterando RD4 para "Em Dúvida + Em Dúvida = Excluído", teríamos o conjunto de regras mais exclusivo sendo aplicado.

Ali e Petersen (2014) realizaram um estudo e concluíram que usar a estratégia mais inclusiva permitiria selecionar todos os estudos relevantes. Contudo, haveria uma sobrecarga (artigos irrelevantes que precisariam ser analisados) de aproximadamente 25% em relação à estratégia mais exclusiva, a qual encontraria 94% dos estudos.

A experiência em pesquisa no tópico do MS é um importante aspecto a ser levado em conta para decidir se um estudo deve ou não ser incluído (KITCHENHAM et al., 2011; WHOLIN et al., 2013). Assim, uma boa alternativa consiste em definir RD3 e RD5 considerando a avaliação do pesquisador mais especialista como sendo aquela a ser

considerada. Contudo, a melhor opção é mesmo buscar atingir um consenso entre os revisores. Quando possível, é muito recomendado ter um terceiro revisor para atuar naqueles casos em que há divergências mais sérias.

Vale reforçar que durante o processo de seleção, palavras-chave que não foram adotadas inicialmente nas buscas podem ser identificadas. Isso pode acontecer por diferentes razões como o pesquisador ser novato na área de pesquisa ou a palavra-chave ser um novo termo dessa área. Se novas palavras-chave forem descobertas, elas devem ser consideradas e desencadearão a reexecução de atividades, a começar pela busca. Conseqüentemente, o processo de seleção também deverá ser reexecutado para os novos estudos. Também é possível que novos campos do formulário de extração de dados sejam revelados durante a leitura dos estudos. Nesse caso, o formulário, elaborado inicialmente durante a criação do protocolo, deve ser atualizado. Alguns revisores, inclusive, realizam uma extração inicial dos dados paralelamente à seleção final, de modo a avaliar o formulário de extração de dados.

Gestão do Conhecimento em Teste de Software (GC-TS)
Processo de Seleção: Na seleção inicial, a doutoranda fez a avaliação de todos os artigos retornados. Toda a amostra foi, então, dividida em duas partes e avaliada por seus dois orientadores. Em casos de divergência, procurou-se discutir e se chegar a um consenso, às vezes incluindo na discussão os três revisores. Casos em que ainda assim persistia a dúvida, os artigos foram levados à etapa seguinte. Na seleção final, o mesmo procedimento foi adotado para a seleção baseada no texto completo.
Resumo da Seleção: Inicialmente foi aplicada a busca automática. Contudo, como apenas poucos estudos foram selecionados (13 dos 440 estudos retornados, já eliminadas as duplicatas), decidiu-se aplicar outras duas técnicas: <i>snowballing</i> e pesquisa direta a autores/grupos de pesquisa. Aplicando-se <i>snowballing</i> , mais 8 estudos foram considerados, sendo que apenas um deles foi selecionado. Aplicando-se a pesquisa direta a autores/grupos de pesquisa, mais 3 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, foram selecionados 15 estudos.

Considerações sobre Avaliação da Qualidade dos Estudos

Uma diferença marcante entre RSs e MSs refere-se à avaliação da qualidade dos estudos. Em RSs, muitas vezes, após a seleção final dos estudos, obtida pela aplicação dos critérios de seleção considerando o texto completo, é realizada uma etapa de 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 informação suficiente disponível para a extração de dados (PETERSEN et al., 2015). Petersen et al. (2015), ao analisarem 52 MSs, identificaram que apenas 14 deles (~27%) fizeram alguma forma de avaliação da qualidade. 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 certo limiar de qualidade. Além disso, caso decida-se avaliar a qualidade dos estudos selecionados, é importante encontrar critérios de qualidade que possam ser aplicados de maneira similar ao longo dos diferentes tipos de estudos identificados (WHOLIN et al., 2013).

5. Extração e Categorização de Dados

Para se obter um MS de qualidade é imprescindível ter um esquema de classificação confiável e bem definido. Além disso, para aumentar a confiabilidade das classificações feitas em um MS, é muito importante que a extração de dados seja revisada por outros pesquisadores e que divergências sejam debatidas, visando atingir o consenso (BRERETON et al. 2007; PETERSEN et al. 2015).

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. A primeira opção é útil para definir esquemas para questões de pesquisa gerais, aplicáveis a vários 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 Engenharia de Software ou o PMBOK para a Gerência de Projetos. Qualquer que seja a abordagem adotada, contudo, é sempre bom consultar especialistas na área. Eles podem até mesmo indicar esquemas de classificação existentes (PETERSEN et al., 2015).

Com base no exposto acima, Petersen et al. (2015) 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.
- 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, recomenda-se utilizar os mesmos esquemas de classificação de (ou similares a) outros MSs. Assim, será possível comparar MSs feitos por grupos diferentes. É possível, p.ex., adotar esquemas de classificação gerais para tipos de pesquisa, veículos de publicação e métodos de pesquisa (PETERSEN et al., 2015).

No que se refere a tipos de pesquisa, a classificação mais adotada é a proposta por Wieringa et al. (2006), a qual considera os seguintes tipos de pesquisa:

- Proposta de Solução (*Solution Proposal*): diz respeito a estudos que discutem novas técnicas (ou revisões de técnicas);
- Pesquisa de Avaliação (*Evaluation Research*): refere-se a estudos que avaliam a prática na indústria;
- Pesquisa de Validação (*Validation Research*): refere-se a estudos avaliando novas técnicas, ainda não implantadas na indústria;
- Artigo Filosófico (*Philosophical Paper*): trata de trabalhos que estruturam um campo de conhecimento de uma nova maneira (p.ex., uma nova taxonomia);
- Artigo de Experiência (*Experience Paper*): refere-se a trabalhos que discutem como alguém fez algo na prática (sem estar aplicando uma proposta de solução claramente colocada).

- Artigo de Opinião (*Opinion Paper*): envolve os artigos que apresentam a opinião de um pesquisador ou grupo de pesquisadores sobre um tópico de pesquisa específico.

Vale destacar que artigos de opinião não se caracterizam bem como estudos primários, pois apenas emitem opinião sobre um tópico de pesquisa. Assim, sugere-se não selecionar este tipo de trabalho. Além disso, um mesmo artigo pode ser classificado em mais de uma dessas categorias, em especial aqueles que apresentam propostas de solução. P.ex., se um trabalho apresenta uma nova solução para um problema e discute uma validação em laboratório, ele deve ser classificado como Proposta de Solução e Pesquisa de Validação.

O esquema de classificação para tipos de pesquisa proposto por Wieringa et al. (2006) não foi desenvolvido para estudos secundários em geral, mas sim para uso na Engenharia de Requisitos. Assim, nem sempre é fácil enquadrar artigos de outras áreas nessas categorias. Isso tem feito com que pesquisadores acabem não aplicando consistentemente esse esquema de classificação e seria indicado revisar este esquema para ajustá-lo para outros contextos (WHOLIN et al., 2013). Petersen et al. (2015) propuseram uma tabela de decisão para ajudar a usar consistentemente o esquema de classificação de Wieringa et al. Neste texto, considera-se uma versão modificada da tabela proposta do Petersen et al. (2015), a qual é apresentada na Tabela 1. As alterações realizadas vão na direção de admitir que um mesmo trabalho seja classificado em mais de uma categoria, em especial, admitir que trabalhos podem apresentar uma proposta de solução e conter uma avaliação (prática ou não). Assim, a Tabela 1 admite que haja trabalhos apresentando apenas propostas de solução (sem nenhuma forma de avaliação), trabalhos apresentando uma proposta de solução e uma avaliação na prática (indústria) e trabalhos apresentando uma proposta de solução e uma validação na academia (laboratório).

Tabela 1 – Classificação de Tipos de Pesquisa

	R1	R2	R3	R4	R5
Condições					
Usado na Prática	-	V	F	V	F
Nova Solução	V	-	-	F	F
Avaliação Empírica	-	V	V	F	F
Framework Conceitual	-	-	-	F	V
Experiência do autor	-	-	-	V	F
Decisões					
Proposta de Solução					
Pesquisa de Avaliação					
Pesquisa de Validação					
Artigo de Experiência					
Artigo Filosófico					

Legenda: V = Verdadeiro; F = Falso; - = irrelevante ou não aplicável; R1-R5: regras.

A principal confusão a respeito desta classificação tem origem na distinção entre pesquisa de avaliação e pesquisa de validação (colunas R2 e R3 na Tabela 1). Note que considerar se a solução sendo validada ou avaliada é nova (ou não) não é um critério para distinguir entre validação e avaliação. Tanto na validação quanto na avaliação, a solução está sendo empiricamente avaliada, entretanto a validação não é feita na prática (i.e., ela é feita em laboratório), enquanto estudos de avaliação acontecem em um contexto industrial real (PETERSEN et al., 2015).

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 co-localizados com ou realizados como eventos satélites de conferências).

Por fim, no tocante à classificação de métodos de pesquisa, Petersen et al. (2015) propõem um esquema de classificação baseado em alguns trabalhos sobre o tema, dentre eles (WOHLIN et al., 2012) e (EASTERBROOK et al., 2008). O esquema de classificação proposto inclui o seguinte conjunto de métodos de pesquisa frequentemente aplicados na Engenharia de Software: *survey*, estudo de caso, experimento controlado, pesquisa ação, etnografia, simulação, prototipagem, e análise e prova matemática.

Petersen et al. (2015) ainda argumentam que a classificação de métodos de pesquisa tem de estar consistente com a classificação do 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. Este 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 ilustra esta relação.

Para permitir uma melhor compreensão dos estudos primários, é muito importante claramente diferenciar entre artigos e estudos. Ainda que pouco usual, um artigo pode conter resultados de vários estudos. Mais comum de acontecer, um estudo pode ser reportado em vários artigos.

Quando um mesmo estudo estiver sendo reportado em mais de um artigo, pode-se escolher o artigo mais recente ou aquele dê uma visão mais abrangente do estudo. Além disso, é importante indicar quais artigos foram excluídos por reportar o mesmo estudo (WHOLIN et al. 2013). Uma solução alternativa é incluir todos os artigos, mas computar, para efeito de extração e agregação de dados, todos eles como um único estudo (KITCHENHAM et al., 2012). Essa solução é mais indicada quando os vários artigos apresentam conteúdos bastante diferentes e reportam, na verdade, diferentes facetas ou até mesmo evoluções do mesmo estudo.

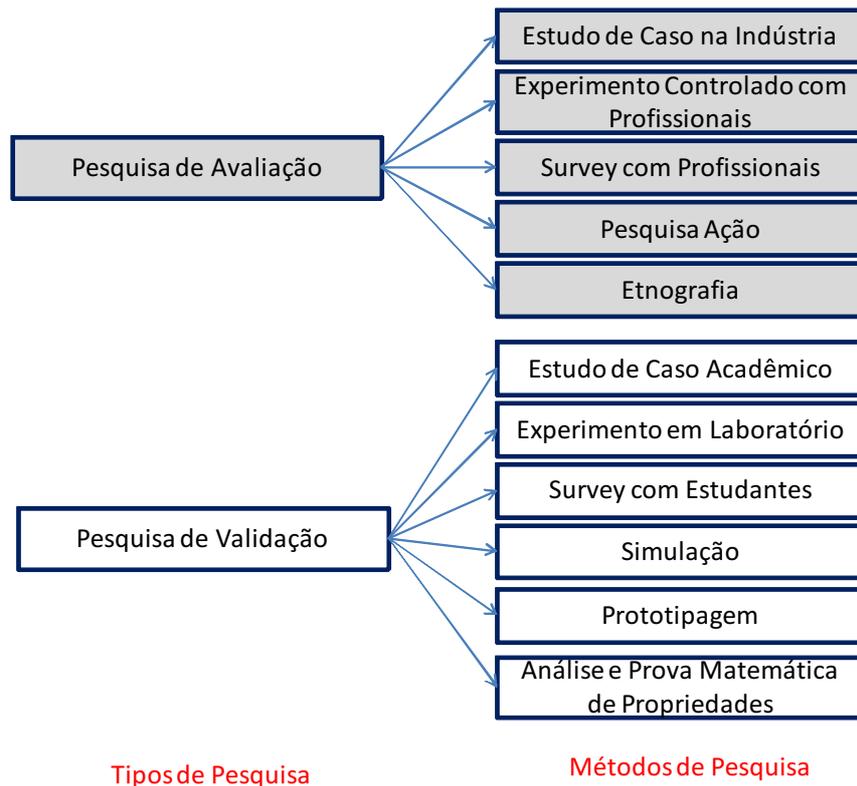


Figura 6 - Relação entre Tipos de Pesquisa e Métodos de Pesquisa (adaptado de (PETERSEN et al., 2015)).

Gestão do Conhecimento em Teste de Software
<p>Para derivar os esquemas de classificação, foram usadas as três abordagens principais discutidas anteriormente: (i) adotar esquemas existentes, (ii) definir um esquema previamente com base na literatura, (iii) deixar os esquemas emergirem dos próprios estudos selecionados. Para as questões de pesquisa QP2, QP3, QP5, QP6, QP8 e QP9, os esquemas de classificação emergiram dos próprios estudos selecionados. Para QP1 e QP4, foram adotados esquemas de classificação existentes (respectivamente os esquemas discutidos anteriormente para veículos de publicação e tipos de pesquisa); finalmente, para QP7, o esquema de classificação foi definido com base na literatura (foi usada a classificação proposta por Nonaka e Takeuchi (1997) para tipos de itens de conhecimento, a qual considera dois tipos de itens de conhecimento: conhecimento tácito e conhecimento explícito), sendo uma categoria posteriormente refinada em um esquema que emergiu dos próprios estudos selecionados (para itens de conhecimento do tipo explícito).</p>

6. Síntese e Apresentação de Resultados

Uma vez que MSs enfocam a categorização de estudos, 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 bem o número de estudos classificados em uma dada categoria. Tabelas também servem bem a esse propósito, sendo possível indicar, ainda, quais trabalhos foram classificados em cada categoria. Contudo, às vezes é interessante combinar informações sobre diferentes categorias. Neste caso, tais ferramentas não são suficientes.

Para agregar dados de diferentes categorias e extrair informações mais complexas, derivadas do cruzamento de duas classificações, 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 no tamanho das bolhas. Assim, é possível representar informações em três dimensões. Quando usados em MSs, o tamanho da bolha é usado para representar o número de estudos que se enquadram em um par de categorias. Assim, gráficos de bolha são muito indicados para ilustrar o número de estudos para uma combinação de classificações (p.ex., subtópico de pesquisa combinado com tipo de pesquisa) (PETERSEN et al., 2015).

Outras formas de visualização usadas em MSs incluem mapas de calor, gráficos de linhas e diagramas de Venn (PETERSEN et al., 2015).

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 desta natureza (PETERSEN et al., 2015). Uma estrutura de artigo indicada inclui as seguintes seções:

- **Introdução:** deve prover um background inicial sobre o tópico estudado e, acima de tudo, uma motivação para a sua importância, de modo a justificar a necessidade de se realizar um MS sobre o assunto. A introdução deve, ainda, procurar destacar a utilidade do MS, podendo também prover uma visão geral de como ele foi conduzido.
- **Background Teórico:** deve prover um background teórico breve, mas mais detalhado, dos tópicos 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 (i.e, outros trabalhos de revisão da literatura, feitos de forma sistemática ou não), se existentes, devem ser brevemente comentados, procurando justificar porque eles não foram considerados suficientes para prover uma visão geral e abrangente do tópico alvo de seu mapeamento. Recomenda-se a realização de um estudo terciário para procurar por trabalhos desse tipo antes de se iniciar os trabalhos de um MS.
- **Método de Pesquisa:** nesta seção devem ser apresentadas as principais informações relacionadas ao método de pesquisa aplicado (no caso, mapeamento sistemático). Assim, devem ser claramente apresentados: (i) o objetivo do MS, (ii) suas questões de pesquisa, (iii) aspectos básicos relacionados à busca e seleção, incluindo: string de busca, fontes pesquisadas, critérios de inclusão, exclusão e de qualidade (estes últimos 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, esta informação pode ser apresentada no final da seção de resultados).
- **Resultados:** esta seção deve iniciar dando uma visão geral do processo de seleção e dos estudos retornados e selecionados em cada etapa 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 aqui. Discussões envolvendo interpretação, análises e conclusões sobre os dados obtidos devem ser colocados na próxima seção.

- Discussão: aqui sim, deve ser feita uma discussão com base na interpretação e análise dos resultados obtidos e apresentados na seção anterior. Tendências e lacunas de pesquisa devem ser apontadas.
- Conclusões: nesta seção deve ser apresentado um resumo do MS e suas conclusões, bem como trabalhos futuros e caminhos a seguir.

Vale destacar que uma publicação reportando os resultados de um MS deve listar todos os artigos selecionados. Um mapeamento que não apresenta a lista de artigos selecionados tem 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 de artigos. Assim, recomenda-se escolher como veículo de publicação para reportar os resultados de um MS um que não imponha muitas restrições. Uma alternativa interessante para evitar este problema consiste em incluir detalhes sobre todos os estudos primários selecionados e suas classificações em um apêndice publicado na Web (KITCHENHAM et al. 2011).

Gestão do Conhecimento em Teste de Software
Foram usados gráficos de barras (para a distribuição dos estudos ao longo do tempo) e várias tabelas, permitindo não apenas apresentar o número (e o percentual) de estudos selecionados classificados em cada categoria, mas também quais estudos foram classificados em cada categoria. Isso foi possível, porque foram selecionados poucos estudos (apenas 15).

7. Ameaças à Validade de um MS

Por melhor que um MS tenha sido planejado e conduzido, sempre haverá algo que representará uma ameaça à validade dos resultados encontrados. Neste sentido, é fundamental identificar ameaças à validade do mapeamento para deixar claro para o leitor o que pode ter comprometido os resultados, dando a ele uma base para definir se seu estudo lhe será útil ou não. Possíveis ameaças à validade de um MS incluem:

- A avaliação dos pesquisadores envolvidos na seleção dos estudos pode ter sido tendenciosa e estudos relevantes descartados. O mesmo pode acontecer em relação às conclusões relativas aos dados coletados, devido a tendências dos pesquisadores envolvidos na interpretação dos dados.
- A amostra de estudos selecionados pode não ser efetivamente representativa do que existe na literatura, havendo diferença significativa entre o que foi selecionado e os estudos existentes que tratam efetivamente do tópico estudado.
- Pode ser difícil generalizar os resultados do mapeamento. Esta ameaça é mais forte quando poucos estudos são selecionados e não há evidências suficientes para se generalizar os resultados.

Referências Bibliográficas

ALI, N.B., PETERSEN, K., Evaluating Strategies for Study Selection in Systematic Literature Studies. In: Proceedings of the 8th International Symposium on Empirical Software Engineering and Measurement, ESEM 2014, 2014.

ARKSEY, H., O'MALLEY, L., Scoping Studies: Towards a Methodological Framework, Int. J. Soc. Res. Meth., 8 (1), 2005, 19–32.

BRERETON, P., KITCHENHAM, B.A., BUDGEN, D., TURNER, M., KHALIL, M., Lessons from Applying the Systematic Literature Review Process within the Software Engineering Domain, *Journal of Systems and Software*, 80 (4), 2007, pp. 571–583.

BUDGEN, D., TURNER, M., BRERETON, P., KITCHENHAM, B., Using Mapping Studies in Software Engineering. In: Proceedings of PPIG 2008, Lancaster University, 2008, pp. 195–204.

DYBA, T., DINGSOYR, T., HANSSSEN, G.K., Applying Systematic Reviews to Diverse Study Types: an Experience Report. In: First International Symposium on Empirical Software Engineering and Measurement, ESEM 2007, 2007, pp. 225–234.

EASTERBROOK, S., SINGER, J., STOREY, M.A., DAMIAN, D., Selecting Empirical Methods for Software Engineering Research. In: Guide to Advanced Empirical Software Engineering, Springer, 2008, pp. 285–311.

KITCHENHAM, B.A., CHARTERS, S., *Guidelines for performing systematic literature reviews in software engineering*. Tech. Rep. EBSE-2007-01, Keele University, 2007.

B. KITCHENHAM, P. BRERETON, AND D. BUDGEN, The Educational Value of Mapping Studies of Software Engineering Literature. In Proceedings of the 32nd International Conference on Software Engineering – ICSE 2010, Cape Town, South Africa, 2010, pp. 589–598.

KITCHENHAM, B., PRETORIUS, R., BUDGEN, D., BRERETON, O.P., TURNER, M., NIAZI, M., LINKMAN, S., Systematic Literature Reviews in Software Engineering – A Tertiary Study, *Information and Software Technology*, vol. 52, 2010, pp. 792–805.

KITCHENHAM, B.A., BRERETON, O.P., BUDGEN, D., Using Mapping Studies as the Basis for Further Research – A Participant-Observer Case Study. *Information and Software Technology*, vol. 53, 2011, pp 638-651.

KITCHENHAM, B.A., BRERETON, O.P., BUDGEN, D., Mapping Study Completeness and Reliability – A Case Study, *IET Seminar Dig.*, vol. 2012, 2012, pp. 126–135.

PETERSEN, K., FELDT, R., MUJTABA, S., MATTSSON, M., Systematic Mapping Studies in Software Engineering. In: Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE 2008), 2008, pp.68-77

PETERSEN K., VAKKALANKA, S., KUZNIARZ, L., Guidelines for Conducting Systematic Mapping Studies in Software Engineering: An Update. *Information and Software Technology*, vol. 64, 2015, pp. 1–18.

SOUZA, E. F., FALBO, R.A., VIJAYKUMAR, N.L., Knowledge Management Initiatives in Software Testing: a Mapping Study. *Information and Software Technology*, v.57, 2015, pp. 378 – 391.

WIERINGA, R., MAIDEN, N. A. M., MEAD, N. R. & ROLLAND, C., Requirements Engineering Paper Classification and Evaluation Criteria: a Proposal and a Discussion. *Requirements Engineering*, 11(1), 102–107.

WOHLIN, C., RUNESON, P., HÖST, M., OHLSSON, M.C., REGNELL, B., WESSLÉN, A., *Experimentation in Software Engineering*, Springer, 2012.

WOHLIN, C., RUNESON, P., SILVEIRA NETO, P.A.M., ENGSTRÖM, E., MACHADO, I.C., ALMEIDA, E.A., On the Reliability of Mapping Studies in Software Engineering, *Journal of Systems and Software*, vol. 86, n. 10, 2013, pp 2594 - 2610.