

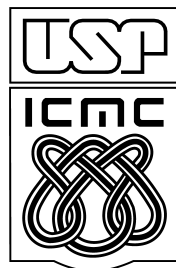
UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Processo de revisão sistemática

Marco Aurélio Graciotto Silva

n. XXX

RELATÓRIO TÉCNICO



São Carlos - SP
Fevereiro/2011

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

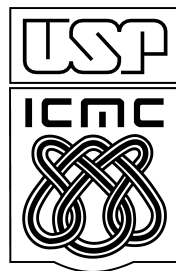
ISSN: 0103-2569

Processo de revisão sistemática

Marco Aurélio Graciotto Silva

n. XXX

RELATÓRIO TÉCNICO



São Carlos - SP
Fevereiro/2011

Prefácio

Este relatório técnico é fruto da execução de algumas revisões sistemáticas, da observação da execução de muitas outras no âmbito do Grupo de Pesquisa em Engenharia de Software do ICMC-USP (São Carlos/SP) e da leitura de artigos de veículos arbitrados e relatórios técnicos dos principais pesquisadores da área.

O objetivo inicial era o estabelecimento de um processo iterativo e incremental para a execução da revisão sistemática. Existia um consenso no grupo de que esta era a forma a ser executada a revisão. De fato, na prática, algumas revisões eram realizadas já com algumas características de tal processo. No entanto, também existe o consenso da comunidade científica de que os riscos à validade dos resultados aumenta com a adoção de tais processos, em especial com a realização de alterações no protocolo. Desnecessário comentar os riscos inerentes de tais modificações (embora este relatório aborde-as em detalhes).

Após uma adequada compreensão desses elementos, deu-se início à redação deste relatório. No entanto, enquanto refletia sobre as diversas revisões, um fato ficou evidente: as técnicas de síntese utilizadas em uma parcela significativa das revisões eram inadequadas e, em alguns casos, ela sequer existia!

A consternação foi ainda maior ao perceber que eu mesmo não tinha conhecimento adequado quanto a essas técnicas. Isso motivou a inclusão de um objetivo adicional neste trabalho: enfatizar a importância da síntese, destacar as técnicas mais adequadas de acordo com as características dos estudos recuperados e das evidências extraídas, e, na medida do possível, descrever a aplicação das técnicas, senão completamente, ao menos o suficiente para que o leitor possa buscar, na literatura, instruções detalhadas para sua aplicação.

Um último desafio às revisões sistemáticas é o volume de dados a serem analisados. Em contraponto ao rigor defendido nos parágrafos anteriores, este relatório advoga a favor da mineração visual de textos para a seleção de estudos e para a análise dos dados. Em especial quanto à seleção, o uso de uma técnica visual é controverso. No entanto, o texto defende a tese de que, embora exista um risco advindo do uso da visualização, esse risco é inferior aquele atualmente assumido quando as expressões de busca são alteradas para “reduzir” a quantidade de estudos recuperados (o que ignora algumas das diretrizes da estratégia PICO).

Não posso dizer que todos os objetivos foram alcançados. De fato, seria muito audaz, em uma área em franco desenvolvimento (Engenharia de Software Experimental), estabelecer um padrão. Falta amadurecimento da área, falta muito mais amadurecimento do autor. No entanto, este texto certamente é uma introdução adequada ao mundo da revisão sistemática e permitirá a análise crítica das revisões já existentes e a execução de novas revisões com resultados mais sólidos.

Gostaria de agradecer a todos os colegas do Laboratório de Engenharia de Software pelas revisões executadas e que serviram de inspiração. Em especial, gostaria de agradecer ao prof. Maldonado, que trouxe essa “novidade”, a revisão sistemática, para nosso grupo e difundiu a cultura de Engenharia de Software Experimental no grupo de pesquisa; à Erika Höhn, pelas aulas e paciência para a revisão das primeiras revisões; ao Fabiano Ferrari e à Viviane Malheiros pelas discussões sobre formas alternativas para a realização de revisões sistemáticas e tantas mais considerações sobre o rigor delas.

Sumário

1	Introdução	1
2	Engenharia de software experimental	3
2.1	Caracterização de um estudo experimental	5
2.1.1	Hipótese	5
2.1.2	População	5
2.1.3	Tratamento	5
2.1.4	Aplicador	5
2.1.5	Variáveis	6
2.1.5.1	Tipos de variáveis	6
2.1.5.2	Escala	6
2.2	Métodos de investigação	7
2.2.1	Levantamentos (surveys)	9
2.2.2	Estudos de caso	9
2.2.3	Experimentos	9
2.2.4	Revisão sistemática	9
2.3	Validade dos estudos	10
2.3.1	Validade de conclusão	10
2.3.2	Validade interna	10
2.3.3	Validade de construção	11
2.3.4	Validade externa	11
2.4	Considerações finais	11
3	Processo	13
3.1	Visão geral do processo	14
3.2	Atores e papéis	16
3.3	Fases	17
3.3.1	Planejamento	17
3.3.2	Seleção de fontes	21
3.3.3	Seleção de estudos	24
3.3.4	Avaliação de qualidade	25

3.3.5	Extração de dados	25
3.3.6	Análise e síntese de dados	25
3.3.7	Empacotamento ou documentação	25
3.4	Considerações quanto ao processo iterativo	25
4	Planejamento	33
4.1	Definição da questão	33
4.2	Definição do controle	36
4.3	Definição dos termos (palavras-chaves)	37
4.4	Construção de expressão genérica de busca	38
4.5	Definição dos critérios de seleção das fontes	39
4.5.1	Acesso	39
4.5.2	Relevância	39
4.5.3	Mecanismos para recuperação de estudos	40
4.5.4	Idioma dos estudos	40
4.5.5	Período indexado	40
4.5.6	Tipos de publicações	40
4.5.7	Tipo de fontes	41
4.5.7.1	Mecanismos de busca	41
4.5.7.2	Pesquisadores	41
4.5.7.3	Publicações	41
4.6	Definição dos critérios de seleção de estudos	41
4.7	Avaliação da qualidade dos estudos primários	43
4.7.1	CASP	44
4.7.2	Impacto	44
4.8	Extração de dados	45
4.9	Análise dos resultados	46
4.10	Considerações finais	47
5	Seleção de fontes	51
5.1	Avaliação do grupo de controle	51
5.2	Seleção das fontes	51
5.3	Avaliação das fontes	53
5.3.1	Caracterização das fontes	54
5.4	Definição de expressões de busca específicas	54
5.4.0.1	ACM Digital Library (antes de 2010)	55
5.4.0.2	CiteSeer Library	56
5.4.0.3	IBM Journal	56
5.4.0.4	IEEE Xplore	56

5.4.0.5	Inspec	58
5.4.0.6	Scirus (Elsevier)	58
5.4.0.7	Compendex	59
5.4.0.8	ScienceDirect	59
5.4.0.9	Biblioteca da Keele University	59
5.4.0.10	Biblioteca Digital Brasileira de Computação	59
5.4.0.11	SpringerLink	59
5.4.0.12	Google Scholar	60
5.4.0.13	ISO	60
5.4.0.14	ISI Web of Knowledge	60
5.4.0.15	JSTOR	61
5.4.1	CiteSeer.IST	61
5.4.1.1	The Collection of Computer Science Bibliographies	61
5.4.2	DBPL	61
5.4.3	io-port.net	62
5.4.3.1	Wiley InterScience	62
6	Seleção de estudos	63
6.1	Procedimentos para identificação de estudos	65
6.2	Procedimentos para a seleção de estudos	65
6.3	Arbitragem da lista e critérios	66
6.3.1	Seleção de estudos	66
6.3.2	Procedimentos para seleção de estudos	66
6.4	CrITÉrios de seleção baseados em VTM	67
7	Extração	69
7.1	Extração de dados básicos dos artigos	70
7.2	Extração de evidências	71
8	Análise	73
8.1	Contagem de voto	74
8.2	Meta-análise	74
8.3	Síntese meta-etnográfica	74
9	Empacotamento	77
10	Conclusões	79
10.1	Ameaças	79
10.1.1	Protocolo	79
10.1.2	Qualidade do experimento	80

10.1.3 Replicação	80
10.2 Documentação	81
10.3 Garantia de qualidade	81
10.4 Trabalhos futuros	81
Referências	82

Lista de Acrônimos

CASP	<i>Critical Appraisal Skills Programme</i>	
ESE	Engenharia de Software Experimental	
GQM	<i>Goal-Question-Metric</i>	1
KDD	<i>Knowledge Discovery and Data Mining</i>	
PICO	Problema ou Popula��o, Interven��o, Compara��o e Resultados esperados ..	35
QIP	<i>Quality Improvement Paradigm</i>	1
VTD	<i>Visual Text Datamining</i>	67
VTM	<i>Visual Text Mining</i>	43

Introdução

A área de Engenharia de Software aproxima-se de seu meio século de existência (mais precisamente 43 anos, contados a partir de 1968) e ainda não se observa a maturidade que outras áreas da Ciência apresentam. De fato, muitos trabalhos ainda não aplicam desenhos metodológicos e técnicas de validação apropriados (ZELKOWITZ; WALLACE, 1998), de modo que a gradual construção e evolução de técnicas de engenharia de software não tem se concretizado (embora, segundo Zelkowitz (2008), esse cenário esteja gradativamente melhorando).

Em relação a esse cenário, a Engenharia de Software Experimental (ESE) se institui para alavancar o uso do método científico na área. As primeiras experiências para aumentar o rigor da pesquisa em Engenharia de Software, com este paradigma metodológico, são atribuídas a Victor R. Basili (BOEHM *et al.*, 2005). Desde 1970, com seus trabalhos sobre *Goal-Question-Metric* (GQM), *Quality Improvement Paradigm* (QIP) e *Experience Factory*, a área tem ganho ferramentas para aprimorar a qualidade de seus trabalhos, gerando evidências confiáveis e transferência de tecnologia para a indústria (SHULL *et al.*, 2006).

No entanto, a obtenção de evidências e a transferência dos resultados das pesquisas para a indústria não ocorre em um único passo. São necessários diversos estudos experimentais, com gradativo aumento de seu rigor (levantamentos, estudos de casos, experimentos) para a formulação de hipóteses sólidas e a obtenção de evidências confiáveis.

Embora a confiabilidade das evidências dependa do tipo de estudo realizado, ela também depende da replicabilidade dos resultados e da síntese de evidências. Resultados obtidos de um único estudo experimental dificilmente são confiáveis (MILLER, 2000). Os motivos são inúmeros: falha na formulação da hipótese (realizar a pergunta errada ou formular a pergunta para uma “resposta” previamente estabelecida, inserido assim um forte viés – além de suscitar questões éticas), erro no desenho do estudo (não identificar variáveis importantes, procedimentos insuficientes para mitigar ou controlar os riscos), erro na execução do estudo (imperícia do experimentador, ocorrência de situações que não estavam previstas no desenho do estudo). Ao mesmo tempo, a simples execução de diversos experimentos, mesmo que relacionados e atrelados a hipóteses com variáveis em comuns ou a mesma hipótese (replicações), não são suficientes para assegurar a qualidade almejada. Logo, é necessária a combinação dos resultados de vários experimentos, de modo que uma única e

confiável conclusão seja obtida (MILLER, 2000).

Dentre os métodos disponíveis em Engenharia de Software Experimental para a síntese de estudos experimentais, uma técnica tem se destacado: a revisão sistemática. Kitchenham transferiu esse instrumento, originário das Ciências Sociais e da Medicina, para a Engenharia de Software, permitindo a identificação, a avaliação crítica e a síntese de evidências de diversos estudos experimentais e o fortalecimento (ou a descrença) de teorias sobre o desenvolvimento de software (DYBA *et al.*, 2007). Adicionalmente, e não menos importante, a revisão sistemática segue um protocolo, estabelecido no primeiro passo da revisão, e é cuidadosamente documentada, permitindo a avaliação crítica e a replicação do estudo. Os mecanismos estabelecidos por esse método de pesquisa propiciam um controle adequado do viés dos estudos primários analisados e do viés inserido pela própria aplicação da revisão sistemática, conseqüentemente reduzindo as chances de se alcançar resultados errados ou equivocados (DYBÄ *et al.*, 2006; MULROW; COOK, 1998; PETTICREW; ROBERTS, 2006).

A técnica de Revisão Sistemática é adotada em diversos grupos de pesquisa, abrangendo desde o ensino da técnica (BALDASSARRE *et al.*, 2008) ao desenvolvimento de projetos de mestrado e doutorado, constituindo-se como um dos primeiros passos do método de pesquisa. As revisões da literatura são um elemento necessário para se manter atualizado quanto às novas evidências que se acumulam nas áreas de interesses dos pesquisadores e para identificar oportunidades de pesquisa (*e.g.*, a ausência de evidências sobre determinado fenômeno) (DYBA *et al.*, 2007). A realização de revisões criteriosas, que contribuam para a correta análise do panorama da área de conhecimento, é um requisito para a construção de hipóteses e, conseqüentemente, de conclusões sólidas. Nesse contexto insere-se o presente relatório técnico: definir um modelo de processo de revisão sistemática voltada para a clara definição da questão e do escopo de pesquisas científicas de pós-graduação. Certamente que o processo também podem ser utilizados para revisões sistemáticas com fins distintos, porém as técnicas de síntese e considerações realizadas neste trabalho são específicas para o panorama de projetos de mestrado e doutorado.

Observando-se as revisões sistemáticas realizadas na área de Engenharia de Software no Programa de Pós-Graduação do ICMC-USP, características e necessidades comuns quanto às revisões foram identificadas: utilização de um processo iterativo, com refinamentos a cada etapa; definição de questões amplas e pouco detalhadas; avaliação da qualidade dos termos utilizados para a busca de estudos primários; utilização dos resumos no estabelecimento dos critérios de seleção de estudos; e emprego de ferramentas para a recuperação dos estudos. Woodall e Brereton (2006) identificaram características semelhantes e concluíram que a mais importante é o processo iterativo, com o refinamento, passo a passo, da revisão. Destaca-se também a importância da realização de incrementos da revisão, mantendo-se as mesmas hipóteses e protocolos previamente estabelecidos, para atualizar a revisão e acompanhar as mudanças do estado da arte durante a realização dos projetos de mestrado e de doutorado.

O restante deste trabalho organiza-se da seguinte forma. Conceitos básicos sobre Engenharia de Software Experimental, necessários não apenas para a realização de revisões sistemáticas, mas para qualquer estudo experimental e, principalmente, para avaliar criticamente as características (e os riscos) de cada estudo, são apresentados no Capítulo 2. O modelo do processo, suas fases e atores são apresentados no Capítulo 3. De fato, esse capítulo define uma visão geral, considerando todas as fases da revisão sistemática. Porém, detalhes quanto a interações no escopo de cada fase são definidos nos capítulos subsequentes: Capítulos 4 e 6 a 8. Tais capítulos especificam as atividades e os procedimentos necessários à execução, inclusive com instruções quanto ao uso de ferramentas de revisão sistemática. Ameaças à validade interna são identificadas e medidas para amenizá-las ou controlá-las são apresentadas para cada fase ou atividade. Finalmente, considerações quanto ao processo e ameaças à validade das revisões sistemáticas como um todo são descritas no Capítulo 10.

Engenharia de software experimental

Um estudo experimental é um ato ou operação com a finalidade de descobrir algo desconhecido ou de provar uma hipótese, envolvendo um investigador coletando dados e executando análises para determinar o que os dados significam (BASILI *et al.*, 1999).

Todo estudo deve especificar, minimamente, os seguintes dados: objetivos, método, resultados e conclusões. Em objetivos, a hipótese de pesquisa deve ser claramente estabelecida. O método estabelece a abordagem utilizada para colher evidências acerca da hipótese a ser aceita ou rejeitada. Os resultados contém não apenas os resultados dos métodos, mas também discutem ameaças a validade dos resultados, observando limitações do método utilizado e das variáveis do estudo.. A conclusão é o desfecho o trabalho, afirmando-se ou não a hipótese e definindo implicações futuras.

O método de pesquisa desempenha um importante papel, estabelecendo o rigor das técnicas e procedimentos adotados e, principalmente, a validade das evidências obtidas do estudo. No entanto, muitos trabalhos em Engenharia de Software carecem de níveis adequados de validação e alguns sequer testam suas hipóteses (ZELKOWITZ; WALLACE, 1998). Apresentado esse cenário, a própria comunidade de Engenharia de Software iniciou uma busca por modelos de pesquisa adequados para a área, atenta a melhorar a qualidade de seus trabalhos.

A pesquisa em Engenharia de Software pode ser realizada sob o paradigma analítico ou científico (GLASS, 1994; BASILI, 1992). Na primeira forma, concentram-se as pesquisas em métodos formais e outras áreas de Engenharia de Software fortemente fundamentadas na Matemática. A segunda fundamenta-se na observação do mundo e na construção de modelos. Ainda no paradigma científico, temos o método de engenharia (no qual as soluções atuais são estudadas, mudanças propostas e avaliadas) e o método experimental (em que modelos são propostos e avaliados por estudos experimentais).

A maioria dos estudos em Engenharia de Software envolve interações com seres humanos em diversos papéis (analistas, projetistas, programadores, testadores, gerentes de projeto, dentre outros) no desenvolvimento de novas técnicas, métodos e ferramentas. Para este tipo de atividade, o método experimental demonstra-se mais apropriado para a obtenção de resultados cientificamente fundamentados (WOHLIN *et al.*, 2000; JURISTO;

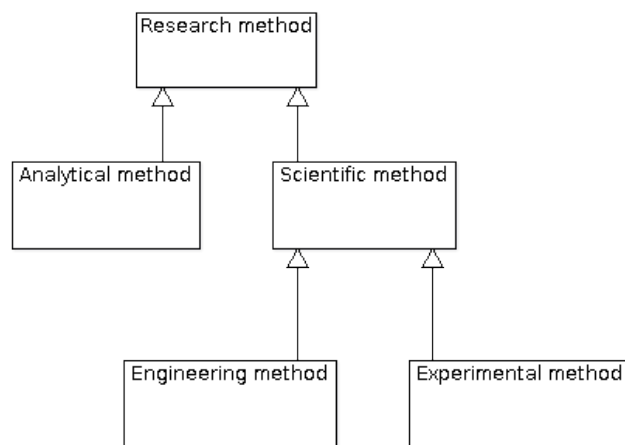


Figura 2.1: Métodos de pesquisa em Engenharia de Software.

MORENO, 2001). Considerando as especificidades da área e as características dos estudos a serem conduzidos, estabeleceu-se a Engenharia de Software Experimental para provar uma sólida base para a pesquisa em Engenharia de Software.

A Engenharia de Software Experimental enfatiza o estudo de Engenharia de Software utilizando princípios científicos para a validação. A partir da formulação de uma teoria, realizam-se estudos que permitam a obtenção de evidências que comprovem a teoria proposta (JURISTO; MORENO, 2001). Esses estudos experimentais determinam tratamentos e resultados, os quais correspondem, respectivamente, a causas e efeitos propostos pela teoria em estudo (Figura 2.2).

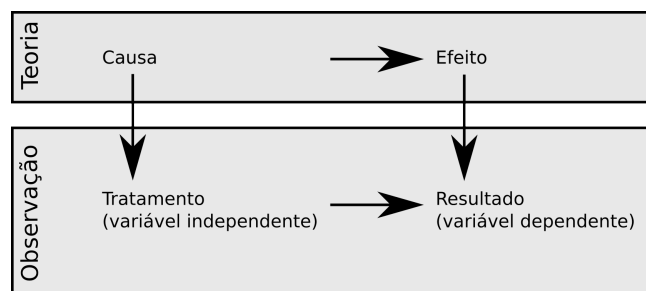


Figura 2.2: Representação dos principais conceitos do paradigma experimental.

De acordo com o tipo dos valores atribuídos às variáveis e obtidos como resultado do trabalho, os estudos experimentais podem ser classificados em qualitativos ou quantitativos (WOHLIN *et al.*, 2000). As pesquisas qualitativas analisam seus objetos de estudo em sua configuração (habitat) natural e interpreta um fenômeno baseado nas explicações que as pessoas formulam quanto a ele. As pesquisas quantitativas centram-se na quantificação de uma relação entre objetos ou na comparação de dois ou mais grupos, buscando a identificação de uma relação causa-efeito (WOHLIN *et al.*, 2000). As duas abordagens são complementares: a quantitativa é apropriada para testar os efeitos de um tratamento, enquanto estudos qualitativos são apropriados para explicar os resultados de uma investigação quantitativa (WOHLIN *et al.*, 2000). De fato, alguns estudos experimentais são uma combinação desses paradigmas (*e.g.*, estudos de caso e levantamentos de dados).

2.1. Caracterização de um estudo experimental

Independente de seu tipo ou desenho, todo estudo experimental possui um conjunto comum de atributos: hipótese, população, tratamento, aplicador, variáveis e ameaças à validade.

2.1.1. Hipótese

A hipótese define o comportamento esperado em função de uma teoria ou suposição que pode explicar um determinado comportamento de interesse de pesquisa. A hipótese nula (H_0) é a hipótese de validação do experimento. Ao refutá-la, consegue-se garantir as condições necessárias para analisar o experimento e avaliar as hipóteses alternativas.

2.1.2. População

A população de um estudo experimental em Engenharia de Software abrange um subconjunto dos componentes básicos de um projeto de desenvolvimento de software: pessoas, produtos, problemas e processos (JURISTO; MORENO, 2001, p. 65).

- **Pessoas:** Desenvolvedores, usuários e demais interessados (*stakeholders*) do projeto de software.
- **Produtos:** Sistema de software e artefatos intermediários (documento de requisitos, código-fonte, desenhos, etc).
- **Problema:** Necessidade ou requisito apresentado pelo usuário ou qualquer outra coisa que tenha originado ou motivado a execução do projeto.
- **Processo:** Conjunto de atividades e métodos utilizados para o desenvolvimento do software.

Dessa população, uma amostra é escolhida para a realização do estudo. Cada elemento escolhido é denominado de unidade do estudo (*experimental unit*). A aplicação de uma combinação de intervenções/-tratamentos em um elemento da amostra da população pelo aplicador do estudo experimental constitui um experimento unitário (JURISTO; MORENO, 2001, p. 62).

2.1.3. Tratamento

O tratamento ou intervenção é a ação a ser aplicada em um objeto do estudo experimental. O tratamento é especificado por um conjunto de variáveis controladas de acordo com o planejamento do experimento: tais variáveis são chamadas de fatores. Os possíveis valores de um fator durante um estudo experimental são chamados de níveis. Logo, o tratamento define os fatores e os níveis que serão aplicados para um elemento da amostra da população.

2.1.4. Aplicador

A pessoa que realiza a intervenção (aplicação de um método, por exemplo) na unidade experimental é denominada de aplicador (*experimental subject*). A forma com que ela aplica a intervenção pode influenciar nos resultados do experimento e, especialmente em Engenharia de Software, ela deve ser apropriadamente caracterizada (JURISTO; MORENO, 2001, p. 58).

2.1.5. Variáveis

As características controladas e medidas em um estudo experimental são as variáveis. Elas são caracterizadas em função de seu tipo e escala.

2.1.5.1. Tipos de variáveis

Quanto ao tipo, as variáveis podem ser um parâmetro, uma variável de bloqueio (*blocking variable*), um fator ou uma resposta:

- **Parâmetro:** Uma variável é um parâmetro se ela não puder ser alterada de dentro do projeto de software, ou seja, ela é uma variável pré-definida. Exemplos de parâmetros são: o problema (necessidade do usuário), informações e outros sistemas utilizados pelo projeto de software (mas que não são definidos e implementados pelo projeto), usuários do software. O conjunto de parâmetros define o contexto do estudo experimental.
Um parâmetro é uma característica estabelecida antes do início do estudo e que deve permanecer inalterada durante a execução dele, de modo que ele não influencie o resultado do estudo experimental. Consequentemente, as evidências obtidas do estudo experimental são válidas para as condições estabelecidas por esses parâmetros (JURISTO; MORENO, 2001, p. 59).
- **Variável bloqueada:** Uma variável de bloqueio ou variação indesejada é uma variável que pode afetar o resultado da intervenção, mas que não pode ser controlada (JURISTO; MORENO, 2001, p. 60).
- **Fator:** Um fator é uma característica do desenvolvimento de software sob estudo a qual afeta a variável de resposta (JURISTO; MORENO, 2001, p. 60). Ele é a variável que será alterada para avaliar o resultado da intervenção sob a população sob estudo. Alguns sinônimos utilizados para fator são: variáveis preditoras, variáveis independentes.
- **Variável de resposta:** São variáveis cuja medida podem ser obtidas apenas após o término do projeto/experimento. Elas representam o efeito dos diferentes fatores aplicados à unidade sob experimentação. A variável de resposta também é denominada de variável dependente.

A identificação das variáveis de um experimento pode ser realizada considerando-se as perspectivas externa e interna. Na perspectiva externa, enxerga-se o projeto sob estudo como uma caixa preta. Ou seja, toda variável que identifica-se no exterior e que, portanto não pode ser alterada de dentro do projeto, é um parâmetro. Já sob a perspectiva interna, observa-se o projeto como se ele fosse uma caixa branca ou aberta, ou seja, é possível observar todos os elementos do projeto. Sob essa óptica, observam-se variáveis que se alteram no início e durante o projeto (variável bloqueada e fator) e variáveis que são lidas apenas ao término do projeto (variáveis de resposta).

2.1.5.2. Escala

Quanto à escala, atribuí-se um valor ou medida a cada variável, seja por definição do aplicador (como parte do projeto do experimento) ou por medições obtidas dos objetos sob estudo. As medidas podem ser tanto em questão a um atributo intrínseco do objeto de estudo (*i.e.*, um atributo interno) ou em relação a um atributo que representa como o objeto em estudo se relaciona com outros objetos (*i.e.*, um atributo externo) (FENTON, 1994). Por exemplo, o tamanho em linhas de código de uma classe é uma medida direta enquanto que o custo para a escrita dessa classe é uma medida indireta.

Toda medida deve ser válida, ou seja, não pode violar quaisquer propriedade do atributo por ela mensurado e também deve fornecer uma caracterização matemática apropriada do atributo (WOHLIN *et al.*,

2000), ou seja, a medida deve preservar a noção intuitiva sobre o atributo e a maneira pela qual ela distingue um objeto de outro (KITCHENHAM *et al.*, 1995).

Uma medida é um número ou símbolo designado para caracterizar um atributo de uma entidade, número ou símbolo este definido pelo mapeamento do mundo empírico (humano, real) para o mundo formal (matemático) (FENTON; PFLEEGER, 1998). Todo mapeamento segue uma escala de medida (*e.g.*, metro, quilograma). Quando medidas feitas em uma escala podem ser transformadas para outra escala, preservando-se os relacionamentos entre os objetos, é dito que tal transformação é admissível (FENTON; PFLEEGER, 1998). Se afirmações realizadas sobre um objeto em uma escala X, após a aplicação uma transformação admissível para uma escala Y, continuarem válidas, é dito que tais afirmações são significativas (*meaningful*), caso contrários elas são sem significado (*meaningless*) (BRIAND *et al.*, 1996). Os tipos de escala mais comuns são:

- **Nominal:** Mapeia um atributo de uma entidade para um número ou símbolo. Esse mapeamento é o mais simples que existe e pode ser interpretado como uma classificação de entidades de acordo com seu atributo (WOHLIN *et al.*, 2000). As transformações possíveis são apenas aquelas que preservam o mapeamento um para um dos objetos.
- **Escalar:** Realiza o mapeamento da entidade para um símbolo ou número, respeitando um critério de ordenação. As transformações devem preservar a ordem das entidades (maior, igual, menor).
- **Ordinal:** Além de ordenada, a diferença entre dois valores possui significado (distância relativa entre duas entidades), embora o valor em específico não. Possui zero arbitrário (*e.g.*, 0°C Celsius).
- **Razão:** Possui zero absoluto (zero absoluto é a ausência, *e.g.*, 0 Kelvin, o qual é a ausência de temperatura).

Quanto a mecanismo de mensuração, uma medida pode ser classificada em objetivo e subjetiva. A medida objetiva é aquela em que não é necessário julgamento para estabelecer o valor da medida, ou seja, o valor depende apenas do objeto sendo medido (*i.e.*, uma medida objetiva terá sempre o mesmo valor, independente de quantas vezes for realizada a medição). A medida é subjetiva se o valor da medida depende do julgamento da pessoa ou do instrumento que realiza a medida.

Medidas que podem ser realizadas sem influência de outras medidas ou atributos, ou seja, que podem ser diretamente obtidas de uma entidade, são denominadas de medidas diretas. Quando a medida depende da mensuração de outros atributos, ou seja, quanto ela é derivada de outras medidas, a medida é chamada de indireta.

2.2. Métodos de investigação

Cada estudo experimental gera evidências que permitem a criação e o aumento da credibilidade e efetividade de teorias quanto a um determinado fenômeno (por exemplo, quanto ao desenvolvimento de software).

A Engenharia de Software Experimental faz uso de dois tipos de estudo: estudos primários e estudos secundários (MAFRA; TRAVASSOS, 2006). Os estudos primários visam a caracterização de uma determinada tecnologia em uso dentro de um contexto específico. Os estudos secundários visam a identificação, a avaliação e a interpretação dos resultados relevantes a um determinado tópico de pesquisa ou fenômeno de interesse (KITCHENHAM, 2004). Dessa forma, os resultados obtidos por diversos estudos primários correlatos atuam como fonte de informação a serem investigadas por estudos secundários (MAFRA; TRAVASSOS, 2006). A precisão e a confiabilidade proporcionadas pelos estudos secundários contribuem para a melhoria e a para o direcionamento de novos tópicos de pesquisa, a serem investigados por estudos primários, em um ciclo iterativo (MAFRA; TRAVASSOS, 2006) (e estudos secundários não podem ser considerados uma abordagem alternativa para a produção primária de evidências (BIOLCHINI *et al.*, 2005)).

Segundo Wohlin *et al.* (2000), os principais tipos de estudos experimentais são: levantamentos, estudos de casos e experimentos. Todos esses são estudos primários. Acrescenta-se ainda à lista a revisão sistemática, um tipo de estudo secundário, que sintetiza as evidências geradas por estudos primários.

Para cada tipo de estudo, é possível definir métodos para sua execução. Cada método caracteriza a questão investigada (geralmente definindo as populações em estudo, a intervenção, o controle, os resultados esperados) e mecanismos para controle das variáveis do estudo.

O grau de confiança que pode ser obtido de um estudo experimental depende diretamente do grau de controle experimental que o pesquisa pode aplicar sobre o objeto sob estudo (WARREN; MOSTELLER, 1993). Dessa forma, é possível estabelecer uma classificação dos tipos de métodos quanto ao grau de confiança de seus evidências.

Embora seja possível estabelecer tal classificação, é necessário esclarecer que todos os métodos são importantes, dado que cada um deles possui um propósito diferente. As evidências com menor grau de confiança fornecem dados necessários para o planejamentos de estudos experimentais com métodos mais rigorosos. Estudos de síntese (tais como revisões sistemáticas) podem fornecer novos elementos a serem estudados por estudos primários, geralmente necessitando de métodos mais simples. Dessa forma, constitui-se um ciclo entre os diferentes métodos de pesquisa experimental.

A escolha do tipo de estudo experimental deve ser realizada de acordo com a questão da pesquisa sendo conduzida. Entretanto, é possível que algumas características do tipo de aplicação sob estudo ou limitações quanto aos recursos disponíveis para a execução do estudo restrinjam o tipo de estudo aplicável. Na Tabela 2.1, alguns fatores e a necessidade ou não daquele fator pelo tipo de estudo são especificados.

Tabela 2.1: Características dos tipos de pesquisa (WOHLIN *et al.*, 2000).

Fator/Tipo	Levantamento	Estudo de caso	Experimento
Controle da execução	Não	Não	Sim
Controle da medição	Não	Sim	Sim
Custo	Baixo	Médio	Alto
Replicabilidade	Alto	Baixo	Alto
Quantitativo	Sim	Sim	Sim
Qualitativo	Sim	Sim	Não

O controle da execução descreve o controle que o pesquisador possui sob o estudo. Em um levantamento, não existe garantia que os objetos de estudo (pessoas) responderão aos questionários ou entrevistas. O mesmo se aplica a estudos de caso, dado que uma empresa pode interromper a aplicação desse a qualquer momento. Um experimento, no entanto, é executado em um ambiente controlado (*e.g.*, em um laboratório) no qual o pesquisador possui autonomia em sua execução.

O controle de medição é o grau com que o pesquisador pode decidir sobre as medidas que serão coletadas.

O custo de investigação é um reflexo do grau de controle proporcionado por cada tipo de pesquisa. Geralmente, experimentos possui um custo de aplicação alto. No entanto, estudos de caso e levantamentos, dependendo do seu tamanho, também podem ter custos elevados (embora, considerando-se a quantidade de objetos de pesquisa, o custo de um levantamento é inferior ao de um caso de uso, que por sua vez possui um custo inferior a um experimento).

A replicabilidade indica o quão replicável é um estudo experimental. Em uma replicação, o estudo é repetido, sob as mesmas condições do estudo original, mas utilizando-se uma população diferente. Se a amostragem foi realizada corretamente (aleatoriamente e representativa da população), espera-se que os resultados da replicação estejam de acordo com os resultados originais. Caso isso não se observe, alguma outra variável, que afeta o resultado do estudo experimental, não foi identificado no desenho original do

estudo e, portanto, deve ser investigada.

Levantamentos de dados e estudos de casos podem ser tanto quantitativos como qualitativos. Experimentos, por outro lado, são exclusivamente quantitativos. Em termos gerais, apenas elementos quantizados podem ser devidamente controlados. De fato, levantamentos não possuem elementos de controle, o que impede a manipulação das variáveis (pode-se argumentar que não existem variáveis independentes em levantamentos de dados). Os estudos de caso possuem variáveis independentes (fatores) e sua atividade é documentada: isto permite algum nível de controle e limitada manipulação dos fatores (no mínimo, o objeto de estudo é o objeto de controle, e este pode ser escolhido especificamente para o estudo). No entanto, não existe um grupo de controle que permita, ao mesmo tempo, alterar os fatores e conseguir distinguir os resultados que são decorrentes do tratamento e os que não são. Em experimentos, define-se a variável de estado, a qual assume diferentes valores e cujo objetivo é permitir a distinção entre uma situação ou grupo de controle e o grupo em tratamento (PFLEEGER, 1995).

2.2.1. Levantamentos (surveys)

Um levantamento é uma investigação realizada em retrospecto, quando, por exemplo, uma determinada tecnologia tem sido utilizada durante um certo período de tempo (PFLEEGER, 1995). Os principais meios utilizados para a coleta de dados são entrevistas e questionários realizados com uma amostra representativa da população sob estudo (MAFRA; TRAVASSOS, 2006). Os resultados são analisados de forma a se extrair conclusões que possam ser generalizadas quanto a população da qual a amostra foi tomada (MAFRA; TRAVASSOS, 2006).

2.2.2. Estudos de caso

Os estudos de casos investigam uma entidade ou fenômeno dentro de um espaço de tempo específico. Durante a condução de um estudo de caso, dados são coletados e, baseado neles, análises estatísticas são conduzidas de forma a se permitir a avaliação de um determinado atributo ou o relacionamento entre diferentes atributos (MAFRA; TRAVASSOS, 2006).

2.2.3. Experimentos

Um experimento é um estudo experimental em que as variáveis envolvidas são manipuladas de forma direta, sistemática e precisa (MAFRA; TRAVASSOS, 2006). O efeito da manipulação das variáveis é medido e, baseado nessas medições, análises estatísticas são conduzidas (MAFRA; TRAVASSOS, 2006).

Em um experimento, dada a possibilidade de controle das variáveis, é possível atribuir tratamentos distintos aos elementos sob estudo (amostra da população), formando grupos de controle. Esse mecanismo permite determinar se o efeito observado é realmente resultado da intervenção aplicada (JURISTO; MORENO, 2001, p. 65).

Em um experimento, as variáveis de resposta estão relacionadas com o conceito de métricas de desenvolvimento de software. Uma variável de resposta tem de ser, necessariamente, quantitativa e medir resultados obtidos após o término do projeto (o que é exatamente o que uma métrica faz).

2.2.4. Revisão sistemática

A revisão sistemática é um método de pesquisa baseado no paradigma experimental que objetiva a obtenção e avaliação das evidências relativas a um tópico de pesquisa em específico (BIOLCHINI *et al.*, 2005). Seu

desenvolvimento, como o próprio nome denota, é realizado de forma sistemática, com uma sequência de passos fixos e executados de acordo com um protocolo definido *a priori* (no início da revisão sistemática). Posteriormente, o protocolo pode ser utilizado por outros pesquisadores, de modo a permitir a replicação do estudo.

Uma vantagem a ser destacada das revisões sistemáticas é que, além do rigor do processo de revisão sistemática (em grande parte advinda da definição do protocolo *a priori*), é que ela permite avaliar hipóteses que não eram alvos dos estudos primários, mas que, com os dados desses e a síntese de seus resultados, torna possível a avaliação da nova hipótese.

No entanto, ela também possui algumas limitações. As técnicas de síntese aplicáveis (e a confiabilidade dos resultados) dependem do tipo de estudo primário utilizado (BIOLCHINI *et al.*, 2005). Também existe o risco da questão da revisão sistemática ser definida de acordo com conhecimento prévio dos pesquisadores quanto a estudos experimentais relativos ao foco da questão. As evidências desses estudos prévios não podem ser utilizada para testar a hipótese/questão. Caso contrário, ocorreria um círculo vicioso entre a hipótese e as evidências (dado que ambas foram obtidas do mesmo conjunto de estudos). Logo, os dados utilizados para gerar a questão da revisão sistemática e os dados utilizados para responder a questão devem ser obtidos independentemente (BIOLCHINI *et al.*, 2005).

2.3. Validade dos estudos

Os resultados de um estudo devem considerar quatro tipos de validade: conclusão, interna, externa e de construção (COOK; CAMPBELL, 1979):

- Validade da conclusão: Validade da análise estatística.
- Validade interna: Validade do relacionamento entre o tratamento e o resultado obtido.
- Validade externa: Validade da generalização dos resultados para uma população maior.
- Validade de construção: Validade da relação entre a teoria e a observação.

2.3.1. Validade de conclusão

A validade de conclusão é relacionada à habilidade de chegar a uma conclusão correta a respeito dos relacionamentos entre o tratamento e o resultado do estudo experimental (TRAVASSOS *et al.*, 2002). Os seguintes aspectos devem ser analisados: teste estatístico, tamanho do conjunto de participantes, confiabilidade das medidas, confiabilidade da implementação dos tratamentos.

2.3.2. Validade interna

A validade interna trata da validade do relacionamento entre o tratamento e o resultado obtido, verificando se ele é causal (e não resultado da influência de algum outro fator que não foi controlado ou medido). Dois tipos de erros devem ser medidos quanto a esse relacionamento:

- Erro do tipo 1: Acontece quando o teste estatístico indica o relacionamento mesmo que não exista nenhum relacionamento real. Ele é calculado pela probabilidade com que a hipótese nula seja rejeitada:

$$P_{typeI\ error} = P\left(\frac{H_0\ rejeitada}{H_0\ verdadeira}\right) \quad (2.1)$$

- Erro do tipo 2: Acontece quando o teste estatístico não indica o relacionamento mesmo que efetivamente exista um relacionamento. Ele é calculado em função da não rejeição da hipótese nula:

$$P_{typeII\text{error}} = P\left(\frac{H_0 \text{ não é rejeitada}}{H_0 \text{ é falsa}}\right) \quad (2.2)$$

Durante a avaliação da validade interna, deve-se prestar atenção aos participantes, à maneira da divisão nas classes, ao modo da aplicação dos tratamentos e aos aspectos sociais.

2.3.3. Validade de construção

A validade de construção considera os relacionamentos entre a teoria e a observação, ou seja, se o tratamento reflete a causa (efeito positivo) e o resultado reflete o efeito positivo. Uma das principais ameaças à validade de construção é o comportamento incorreto dos participantes ou do aplicador do estudo experimental. Alguns exemplos de erros comuns em estudos experimentais: participantes envolvidos podem basear seu comportamento nas suposições sobre as hipóteses; o pesquisador pode projetar o estudo baseado no resultado esperado para o experimento (TRAVASSOS *et al.*, 2002).

2.3.4. Validade externa

A validade externa define as condições que limitam a habilidade de generalizar os resultados de um estudo experimental para a prática industrial. Problemas típicos observáveis: amostra da população não é representativa, instrumentação inadequada para a prática industrial, restrições quanto ao tempo necessário para executar o estudo experimental.

2.4. Considerações finais

Estudos experimentais permitem a obtenção de evidências quanto a um determinado fenômeno ou tratamento sob um determinado contexto, definido pelas variáveis de tipo parâmetro. Suas hipóteses são testadas quanto aos resultados obtidos da aplicação do método e os valores das variáveis de resposta. Riscos são medidos em função do método, de sua aplicação pelo pesquisador ou aplicador da intervenção e das variáveis de bloqueio. A Figura 2.3 fornece um diagrama que representa todos esses elementos de forma sucinta.

Estudos simples, tais como os levantamentos de dados, possuem pouca força para provar uma teoria (SHULL; FELDMANN, 2008, p. 338). Mesmo se considerando um experimento, cujos resultados possuem uma força e qualidade fortes quanto à hipótese trabalhada, os resultados podem não ser suficientes, dado que os resultados obtidos de um único estudo experimental dificilmente são confiáveis (MILLER, 2000). Assim, é necessário acumular os resultados de vários estudos para obter teorias robustas e úteis (BASILI *et al.*, 1999; MILLER, 2000; KITCHENHAM *et al.*, 2004). Uma das técnicas aplicáveis para esse fim é a revisão sistemática, que é a técnica explicada neste relatório técnico.

Cabe esclarecer que não é razoável estabelecer um grau de importância entre os diversos tipos de estudos experimentais. De fato, cada estudo possui finalidades específicas e, muitas vezes, ortogonais. Por exemplo, um experimento controlado sobre o uso de um medicamento para dor (intervenção) em idosos (população) pode aferir o aumento do risco de doenças cardíacas na população investigada. No entanto, os efeitos colaterais de tal medicamento nem sempre são adequadamente representados ou identificados na amostra da população analisada. Um levantamento, que pode ser amplo, é capaz de identificar os casos em que ocorreram efeitos colaterais que, embora raros, podem ser significativos para o tratamento de casos específicos.

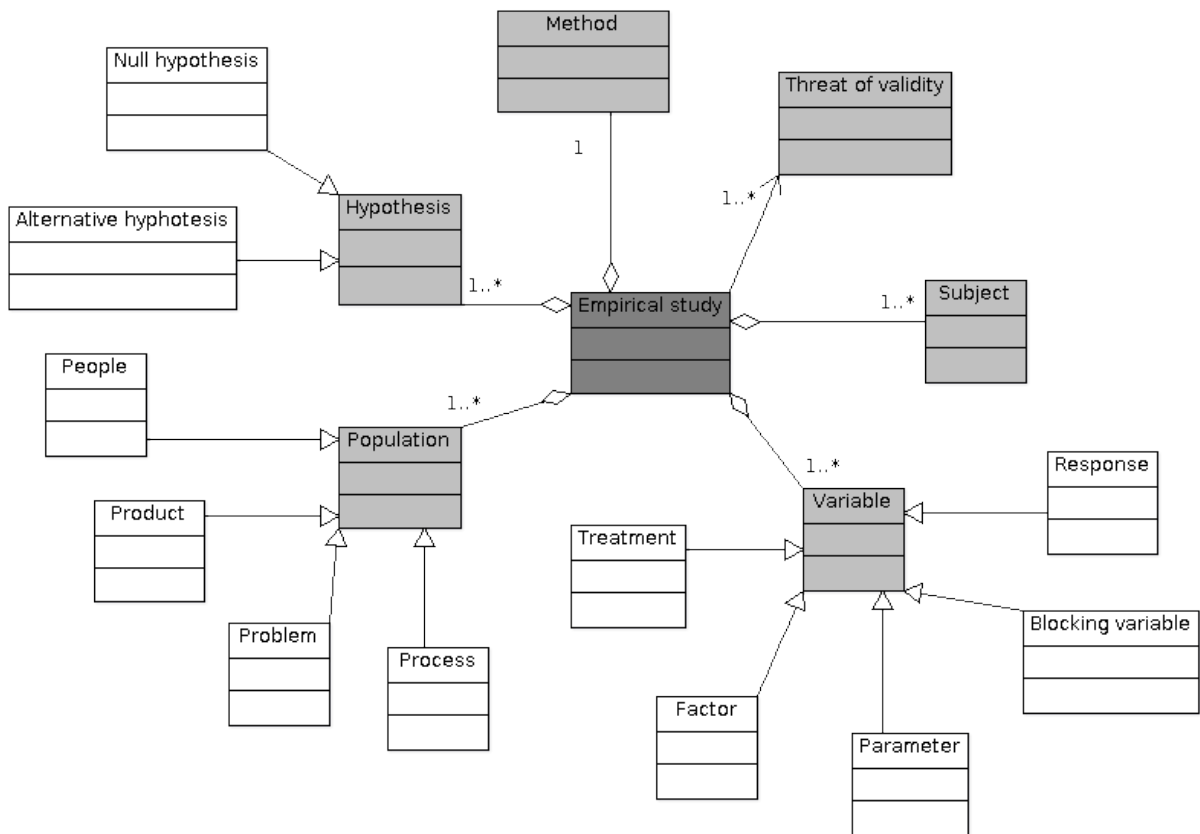


Figura 2.3: Elementos necessários em um estudo experimental.

Finalmente, o objetivo deste capítulo foi a apresentação dos tipos de estudos experimentais geralmente utilizados em Engenharia de Software. No entanto, as informações aqui dispostas não são suficientes para a adequada execução dos tipos de estudos apresentados. Quanto aos levantamentos de dados, uma boa indicação de leitura é o livro de Babbie (1990). Os livros de Yin (2002) e Stake (1995) são recomendados para a execução de estudos de caso. Para experimentos, o livro de Montgomery (2000) é uma boa referência.

Processo

A revisão sistemática é uma metodologia de pesquisa utilizada para identificar, avaliar e interpretar toda a pesquisa relevante disponível sobre uma questão de pesquisa, tópico ou área de interesse específicos (KIT-CHENHAM, 2004). Ela é realizada de maneira formal e sistemática, seguindo uma sequência bem definida de passos especificada em um protocolo estabelecido *a priori* (BIOLCHINI *et al.*, 2005).

Uma recente tendência em projetos de pós-graduação é a utilização de revisões sistemáticas em substituição aos tradicionais levantamentos bibliográficos, tipicamente realizados no início das pesquisas para a caracterização do estado da arte. Dessa forma, assegura-se um maior rigor ao trabalho e se uniformiza essa parte do trabalho científico.

Deve-se observar, entretanto, que uma revisão sistemática tradicional, conforme os preceitos da Engenharia de Software Experimental, promove (1) a avaliação objetiva e a síntese de resultados experimentais de relevância para uma questão de pesquisa particular e (2) a integração daquela evidência na prática profissional (BRERETON *et al.*, 2007). Tratando-se de um revisões sistemáticas utilizadas na caracterização de projeto de pesquisa, o resultado obtido da avaliação objetiva e da síntese constitui o ponto de partida do projeto, ou seja, a justificativa da pesquisa em voga. Ao invés de alcançar a solução de um problema, o objetivo da revisão sistemática é justamente validar a hipótese nula (ou seja, de que para o tópico sendo pesquisado não foi solucionado ou que existem questões em aberto), mas sem responder a pergunta. A resposta dessa, caso existisse, anularia a originalidade da pesquisa. Além dessa consideração acerca do primeiro objetivo da revisão sistemática, o segundo objetivo, a integração da evidência na prática profissional, deve ser ignorado na contexto de projetos de pesquisa com caráter de inovação.

Definida a abrangência da revisão sistemática no contexto proposto, torna-se necessário verificar a executabilidade da técnica. Trata-se de conhecimento notório que revisões sistemáticas são custosas (SHULL; FELDMANN, 2008) e demandam esforços superiores a levantamentos bibliográficos. Desta forma, são tradicionalmente conduzidas por várias pessoas. No entanto, no contexto de um projeto de pesquisa, conduzido por um grupo reduzido, para não dizer uma única pessoa (o pesquisador) ou duas (pesquisador e orientador), seria inviável a condução das atividades requeridas sem considerações específicas e adaptações (WOODALL;

BRERETON, 2006).

O ideal seria que a revisão sistemática fosse executada em uma batelada. No entanto, é comum que erros sejam identificados no protocolo. Logo, é desejável que o processo preveja as alterações e organize o processo de revisão sistemática em iterações de custo adequado, duração previsível e sem perda de qualidade. Um cenário típico seria:

revisão sistemática foi conduzida em três iterações. A primeira iteração foi destinada à criação do protocolo da revisão e a condução de um estudo piloto. Deficiências nos critérios e informações incompletas nos procedimentos foram identificados e corrigidos antes da execução da segunda iteração. Esta executou a revisão, seguindo o protocolo revisado. Alterações significativas no protocolo (alteração de critérios para seleção de fontes) não foram permitidas a partir desse momento e foram consideradas apenas esclarecimentos nos critérios de seleção de estudos. A terceira iteração utilizou o mesmo critério da segunda, mas foi conduzida um ano depois, visando atualizar os resultados anteriormente obtidos com os novos estudos publicados no período.

A

Isto exposto, um processo de revisão sistemática iterativa é definido nas próximas seções. Para cada uma das fases identificadas, um capítulo foi escrito para detalhar as atividades e procedimentos relacionados.

3.1. Visão geral do processo

Revisões sistemáticas são organizadas em sete fases: (1) planejamento, (2) seleção de fontes, (3) seleção de estudos, (4) avaliação da qualidade dos estudos, (5) extração de dados, (6) análise de resultados e (7) documentação da revisão.

- **Planejamento:** Definição dos objetivos da pesquisa e criação do protocolo.
- **Seleção de fontes:** Definição de critérios de seleção (inclusão e exclusão) de fontes de estudos primários.
- **Seleção de estudos:** Definição de critérios de seleção (inclusão e exclusão) de estudos primários.
- **Avaliação:** Avaliação da qualidade dos estudos primários.
- **Extração:** Extração dos dados.
- **Análise:** Síntese dos dados obtidos da execução.
- **Documentação:** Relato da execução (protocolo, bem como as alterações sofridas e decisões tomadas quanto a sua execução) e dos resultados obtidos da revisão sistemática.

O processo segue o modelo iterativo, embora a maioria das revisões sejam executadas em apenas uma iteração (ou seja, no modelo cascata). A fase de documentação é executada continuamente, em paralelo com as demais fases. Entre cada uma das demais fases, são realizadas atividades de garantia de qualidade, de modo a assegurar que os riscos foram devidamente identificados e tratados na fase anterior. Uma vez alcançada uma nova fase, não é possível retornar para a fase anterior.

A literatura relata formatos diferentes deste processo, porém equivalentes à esta apresentada. Brereton *et al.* (2007) definem um processo dividido em três fases: planejamento, condução e documentação. As fases de seleção, extração e análise, do processo descrito neste trabalho, estão inseridas na fase de condução. No trabalho de Biolchini *et al.* (2005), as fases são planejamento, execução, análise e documentação (empacotamento), sendo que as fases de seleção e extração deste trabalho estão inseridas da fase de execução daquele. Finalmente, o processo relatado definido por Kitchenham (2004) contém três fases (planejamento, condução e documentação). A fase de planejamento inclui a identificação da necessidade da revisão sistemática e o desenvolvimento do protocolo, as quais estão incluídas na fase de planejamento do processo aqui definido. As atividades de

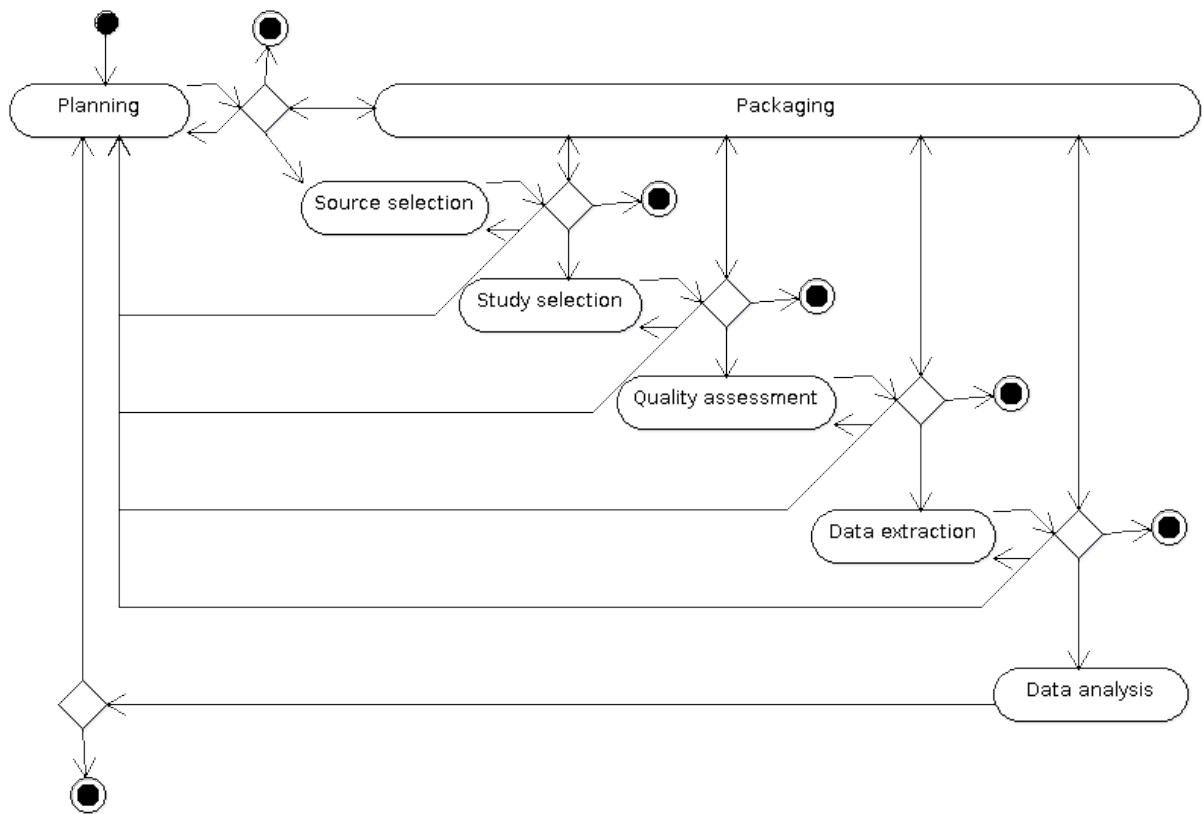


Figura 3.1: Processo de revisão sistemática.

identificação e seleção de estudos, avaliação de qualidade, extração e síntese de dados, pertencentes à fase de condução do processo de Kitchenham (2004), estão definidas, respectivamente, em seleção de estudos, avaliação de qualidade, extração de dados e análise neste processo. Em sua, o processo aqui descrito contempla todos os demais processos citados na literatura, agregando atividades de verificação e validação explícitas entre cada fase. Notadamente, é aceita a condição de falha em qualquer fase do processo, sendo que o sucesso da revisão sistemática é determinado apenas após a execução da análise dos dados (momento em que as questões, definidas no planejamento, são respondidas).

Tipicamente, as fases são executadas sequencialmente. No entanto, a prática tem demonstrado que esta forma de execução não é uma solução adequada. Partes do protocolo, em especial aquelas referentes a critérios de seleção e exclusão, extração de dados e análise, precisam ser revistas e testadas sucessivamente. As práticas tradicionais de revisão sistemática ditam que a execução da revisão inicia-se apenas após a conclusão do protocolo. O raciocínio é que, preservando-o, garante-se que as publicações (e os estudos delas identificados) serão analisadas de forma justa, sob os mesmos critérios. No entanto, caso uma falha no protocolo seja detectada ou, no decorrer da revisão, novas fontes e estudos sejam publicados, essas práticas impedem a inclusão desses estudos, o que, embora não comprometa a validade do estudo, compromete a qualidade dos resultados quanto as pesquisas mais recentes desenvolvidas e avaliadas pela comunidade. Observa-se aqui um contrasenso: a revisão sistemática visa a síntese das evidências de todos os estudos da área sob uma determinada questão; porém a própria revisão sistemática dita que os erros do protocolo, mesmo que afetem os estudos analisados, não devem ser corrigidos (ou melhor, a correção caracteriza uma nova revisão sistemática, sem qualquer relação com a anterior).

Uma questão sempre destacada na revisão sistemática é que ela reúne todas as evidências sobre um determinado tema. Seria muita presunção assumir que todas as evidências são publicadas, assim como que todas as publicações foram analisadas. Toda revisão sistemática assume um risco, calculado, de que algumas evidências não serão obtidas e que alguns estudos, mesmo que relevantes, serão desconsiderados. Logo, embora a revisão tenha um resultado com um grau de confiança elevado, ele não é de 100%. Não permitir correções no protocolo que permitam o aumento da confiança do resultado obtido não resultará em resultados melhores, muito pelo contrário. O único perigo das correções é a redução da confiabilidade e, principalmente, assumir que os resultados das diversas revisões (e a síntese deles) possuem a mesma força. Deve-se sim assumir que são resultados distintos, porém deve-se preservar o fato de que eles são resultados relacionados, que tratam exatamente da mesma questão, população, intervenção, controle e resultados.

A Engenharia de Software encontra um problema semelhante quanto ao desenvolvimento de programas de computador: os requisitos sofrem alterações no decorrer do desenvolvimento e, caso os demais artefatos não sejam alterados a contento, o produto final, apesar de atender os requisitos iniciais, não atende as necessidades dos usuários, que deveria o alvo principal do software. Essa é a situação quando adota-se o modelo sequencial de ciclo de vida de software.

Uma solução para essa questão é a adoção de modelos iterativos (DYBA *et al.*, 2007). De fato, em alguns contextos, revisões sistemáticas com elementos iterativos são desejáveis (OATES; CAPPER, 2009). Com as devidas precauções, iterações podem ser utilizadas para o processo de revisão sistemática. Elas podem ocorrer no contexto da fase em si ou considerando-se a revisão sistemática como um todo (FERRARI; MALDONADO, 2008).

Cabe diferenciar o modelo iterativo do incremental e as ameaças a validade inseridos por cada um deles. Ao realizar as iterações, restritas à uma determinada fase do processo, as ameaças são locais e de tratamento simplificado. No entanto, realizar uma iteração que envolva todas as fases acrescenta riscos à validade que precisam ser cuidadosamente tratados. Por exemplo, alterações no protocolo, entre uma interação e outra, podem prejudicar a análise e síntese dos resultados obtidos em cada interação ou introduzir um viés na revisão. Em contraste, a nova interação pode incluir novos mecanismos para assegurar a qualidade da revisão, corrigindo erros identificados das revisões anteriores. Ainda assim, deve-se adotar um critério sobre a validade dos resultados anteriormente obtidos (e a eventualidade de desconsiderá-los em favor daqueles obtidos nas novas interações). A realização de uma revisão de forma incremental, no entanto, sem alterar o protocolo (*e.g.*, apenas incluindo estudos publicados em períodos posteriores ao anteriormente investigados) não apresenta riscos significativos.

Deve-se considerar também que a realização de diversas revisões sistemáticas sobre uma mesma questão, com graus crescentes de força de seus resultados (seja pela consideração de novos estudos ou pela adoção de novas atividades de verificação e validação que assegurem uma melhor qualidade à revisão em si), podem convergir para conclusões mais fortes.

3.2. Atores e papéis

O ator principal da revisão sistemática é o estudante de mestrado ou doutorado. Ele participa em todas as fases do processo e é o responsável pelo trabalho como um todo. Ele é auxiliado pelo orientador e demais especialistas, que são atores secundários do processo. A responsabilidade desses atores é, principalmente, quanto a revisão do protocolo e dos resultados: sugestão de critérios para a seleção de fontes de estudo e de estudos; e detecção de tendências.

3.3. Fases

Para cada fase, são identificados os papéis, as atividades e os artefatos utilizados e produzidos.

3.3.1. Planejamento

A fase de planejamento consiste em um conjunto de atividades que, a partir de uma problema ou idéia de pesquisa, produz um protocolo para revisão sistemática. As atividades executadas são:

1. Definição da questão.
2. Definição dos atributos da questão.
3. Definição dos critérios de seleção de fontes.
4. Definição dos critérios de seleção de estudos.
5. Definição do grupo inicial de estudos de controle.
6. Definição dos dados a serem extraídos.
7. Definição da técnica de análise e síntese dos dados.

A definição da questão deve ser realizada pelo pesquisador (papel **Researcher**). Tipicamente, ambos os atores, aluno e professor, desempenham esse papel. Com base na literatura científica da área da pesquisa (contendo as limitações atuais das técnicas) e as evidências acerca do objeto de pesquisa, eles possuem idéias para as quais formularão questões a serem respondidas pela revisão sistemática.

A questão é descrita como uma questão simples ou então na forma de hipóteses nula e alternativa. Quando utilizando questões, sempre se define uma ou duas questões principais, que tratam do principal tópico de pesquisa, e várias subquestões, secundárias. As questões secundárias são utilizadas para a extração dos dados, mas não são necessárias para responder a questão primária.

- **Hipótese nula:** Estabelece que não existe relacionamento causal entre os fatores do estudo experimental e os resultados obtidos ou que os fatores não provocaram efeito algum, ou seja, os resultados da intervenção são fato do acaso, coincidências ou não produzem efeitos.

Um exemplo comum de hipótese nula é a de que não existe diferença entre os resultados obtidos pela aplicação da intervenção nos grupos de controle e de tratamento.

- **Hipótese alternativa:** Estabelece que existe uma relação entre os fatores do estudo e os resultados obtidos.

A vantagem da definição da questão utilizando hipóteses nula e alterativa é que é possível estabelecer o nível de significância, ou seja, a probabilidade do resultado ter sido obtido por acaso (e não em decorrência dos fatores do estudo experimental).

A definição dos atributos da questão (população, intervenção, controle e resultados) também é realizada pelos pesquisadores, de forma análoga à definição da questão. A entrada a atividade é a questão elaborada na atividade anterior e os resultados são os atributos da questão: população, intervenção, controle e resultados esperados.

A partir dos atributos da questão, cria-se a expressão de busca a ser utilizada para a recuperação de estudos primários. Além dos termos diretamente identificados da propriedades, também devem ser identificados sinônimos ou termos fortemente relacionados e reconhecidamente adotados na literatura científica. Listas de sinônimos e ontologias podem ser utilizadas para esse fim. Pesquisadores externos podem sugerir também os termos e, principalmente, avaliar se os propostos são relevantes.

Os critérios de seleção de fontes devem permitir a seleção de estudos primários que contemplem os atributos da questão, ou seja, que permitam a seleção de fontes em que sejam recuperáveis estudos sobre a intervenção

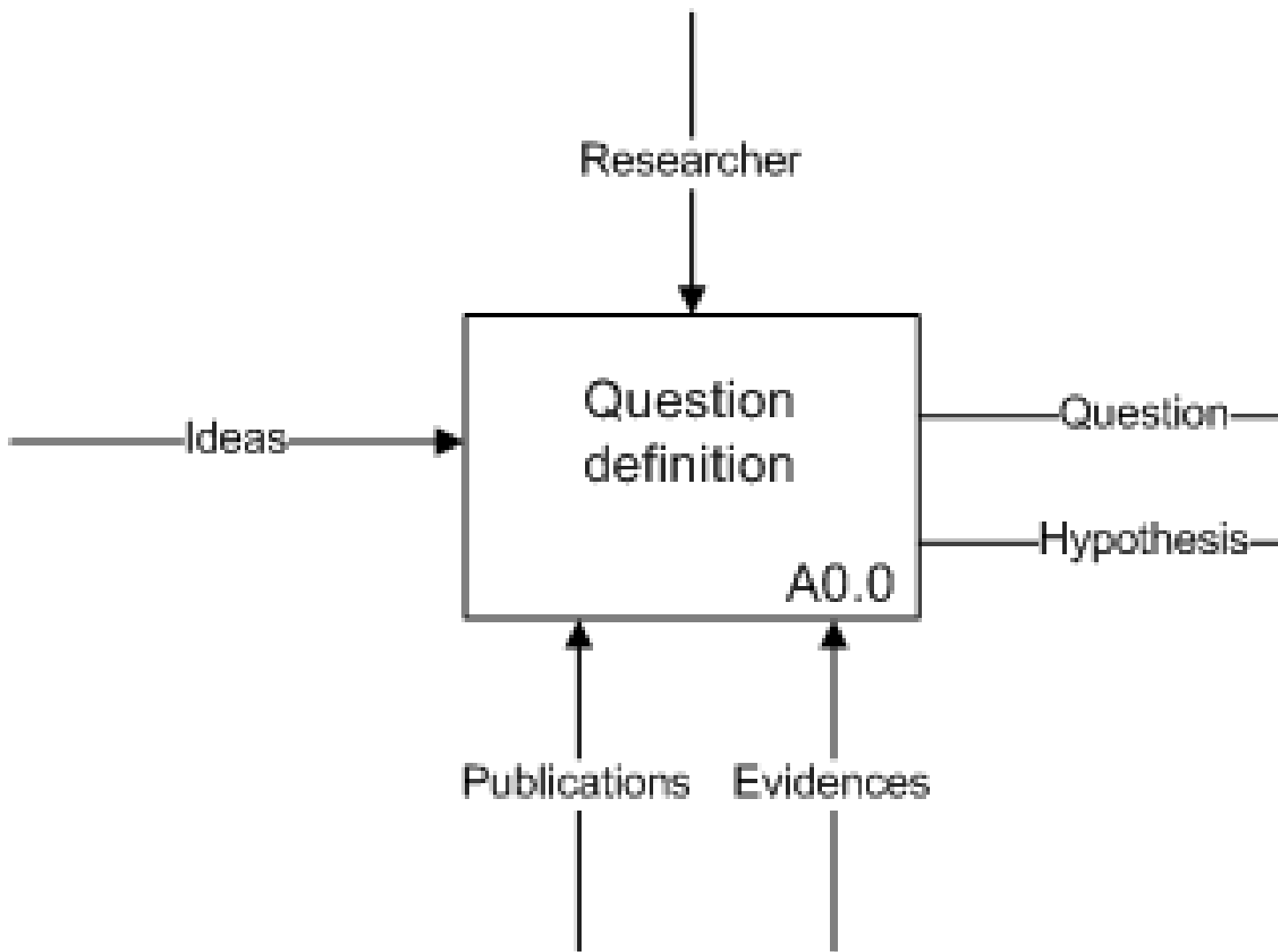


Figura 3.2: Descrição da atividade de definição da questão da revisão sistemática (*Question definition*).

em investigação para uma determinada população. Inicialmente, as fontes de estudo devem contemplar veículos de publicação científica (revistas e eventos arbitrados), mas também é necessário consultar outros pesquisadores (que não sejam os mesmos que conduzem a pesquisa). Algumas áreas definem qualificadores de publicações (por exemplo, no Brasil temos o *Qualis/CAPES*, na Austrália existe o ERA/ARC).

Os critérios de seleção de estudos são definidos a partir da questão e suas propriedades. No entanto, vários dos critérios de seleção são provenientes da prática comum de revisões sistemáticas (por exemplo, restrições quanto ao idioma) ou a força desejável para as conclusões (por exemplo, com base no tipo de estudo aceitável para seleção).

Caso um grupo de controle não tenha sido estabelecido durante a definição das propriedades da questão, é mandatória a definição de um conjunto de estudos a serem utilizados como controle do processo de seleção de fontes e de estudos. Isso é aplicável principalmente para revisões sistemáticas que sejam, de fato, um mapeamento sistemático, ou seja, estudos que visam a identificação e caracterização do estado da arte quando a um assunto, sem necessariamente responder a uma questão focada de pesquisa.

Em mapeamentos sistemáticos, sugere-se a criação de grupos de controles que abranjam não apenas estudos (publicações) que devam estar presentes na revisão sistemática, mas também fontes de estudos. Dessa forma, é possível assegurar que a escolha das fontes e dos estudos tenha sido realizada de forma justa e com

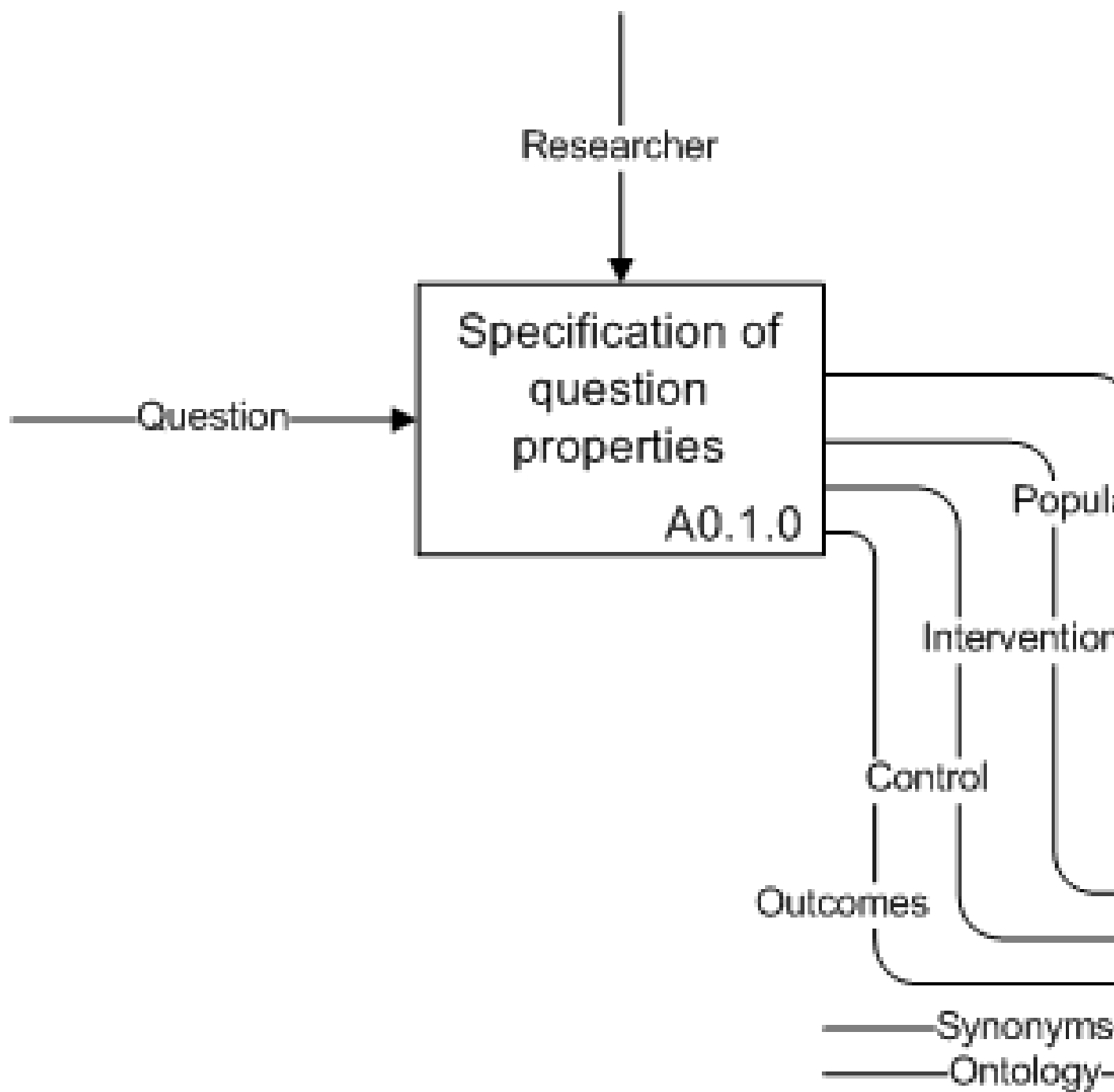


Figura 3.3: Especificação das propriedades da questão da revisão sistemática (*Specification of question properties*).

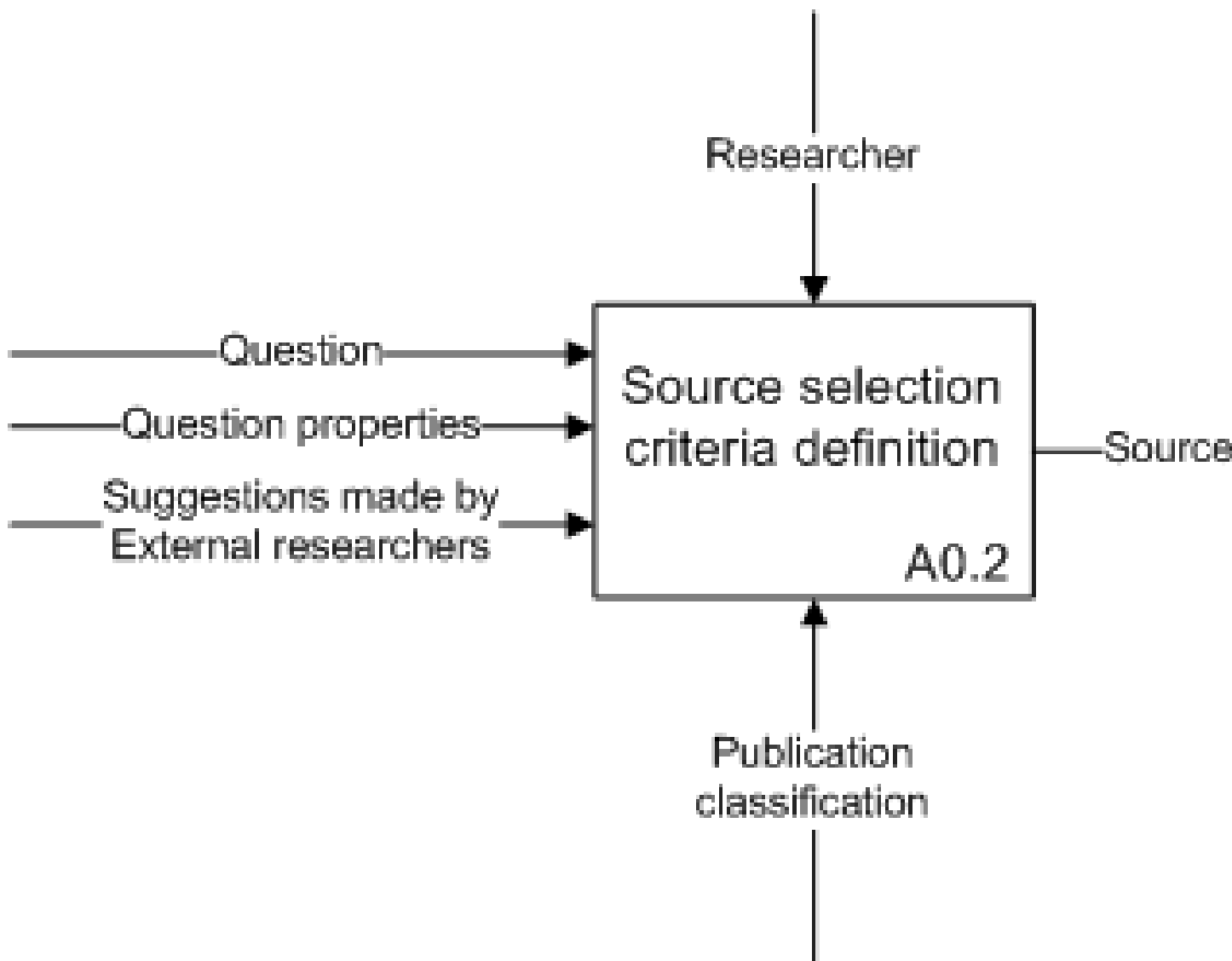


Figura 3.4: Especificação dos critérios de seleção de fontes (*Source selection criteria definition*).

isonomia.

Os dados a serem extraídos dos estudos devem ser definidos a partir das questões definidas no início do protocolo. Para cada questão, os pesquisadores devem definir ao menos um dado que deve ser extraído (portanto, se alguma questão não estiver relacionada a um dado extraível, deve-se avaliar se a questão é realmente relevante para a pesquisa).

Finalmente, os pesquisadores devem escolher qual o método que será utilizado para sintetizar os dados (e as evidências) extraídas dos estudos selecionados. Para tanto, deve-se contar com o auxílio de um estatístico, de modo a definir o método adequado de síntese e quais os requisitos para a correta aplicação desse método.

Cabe também à atividade de definição do método de análise a identificação de ameaças à validade dos resultados obtidos e da revisão sistemática como um todo, considerando todas as suas fases.

De acordo com o método de análise e síntese escolhido para a revisão sistemática, é necessário garantir que os estudos selecionados possuem uma qualidade tal que permita a realização da análise sem o comprometimento dos resultados. Critérios para a avaliação da qualidade dos estudos devem ser definidos para aferir esse atributo.

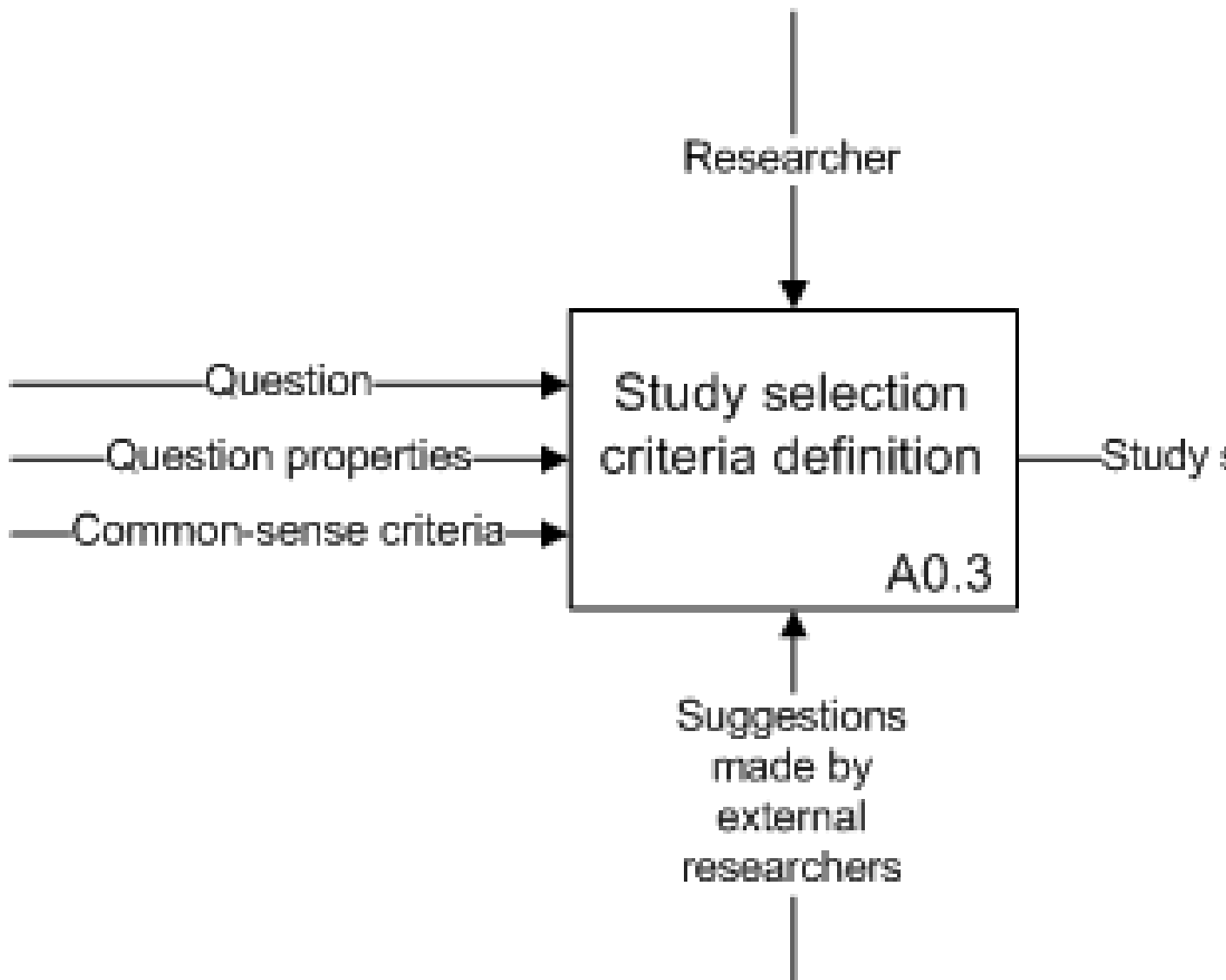


Figura 3.5: Especificação dos critérios de seleção de estudos (*Study selection criteria definition*).

Observa-se que, normalmente, tais critérios não são utilizados para excluir artigos que foram selecionados pelos critérios de seleção de artigos. O objetivo é aferir a força das conclusões obtidas a partir dos estudos selecionados. Caso os estudos encontrados não satisfaçam as condições necessárias para a aplicação do método de análise, um novo método de análise deve ser escolhido ou os critérios de seleção (ou outros elementos do protocolo do qual tais critérios dependam) devem ser alterados.

3.3.2. Seleção de fontes

A seleção de fontes tem início pela aplicação dos critérios de seleção para as fontes definidas no grupo de controle. Em tese, elas devem atender os critérios de inclusão e falhar nos critérios de exclusão. Caso isso não se observe, deve-se proceder da seguinte forma: (1) excluir as fontes que não atenderam os critérios do grupo de controle; ou (2) alterar os critérios para que as fontes possam ser incluídas no grupo de controle.

A primeira alternativa representa a falha dos pesquisadores, internos e externos, em fornecer fontes relevantes. Como a revisão sistemática depende da interpretação dos dados por parte de pessoas (principalmente em Engenharia de Software, área na qual os trabalhos não relatam de forma explícita

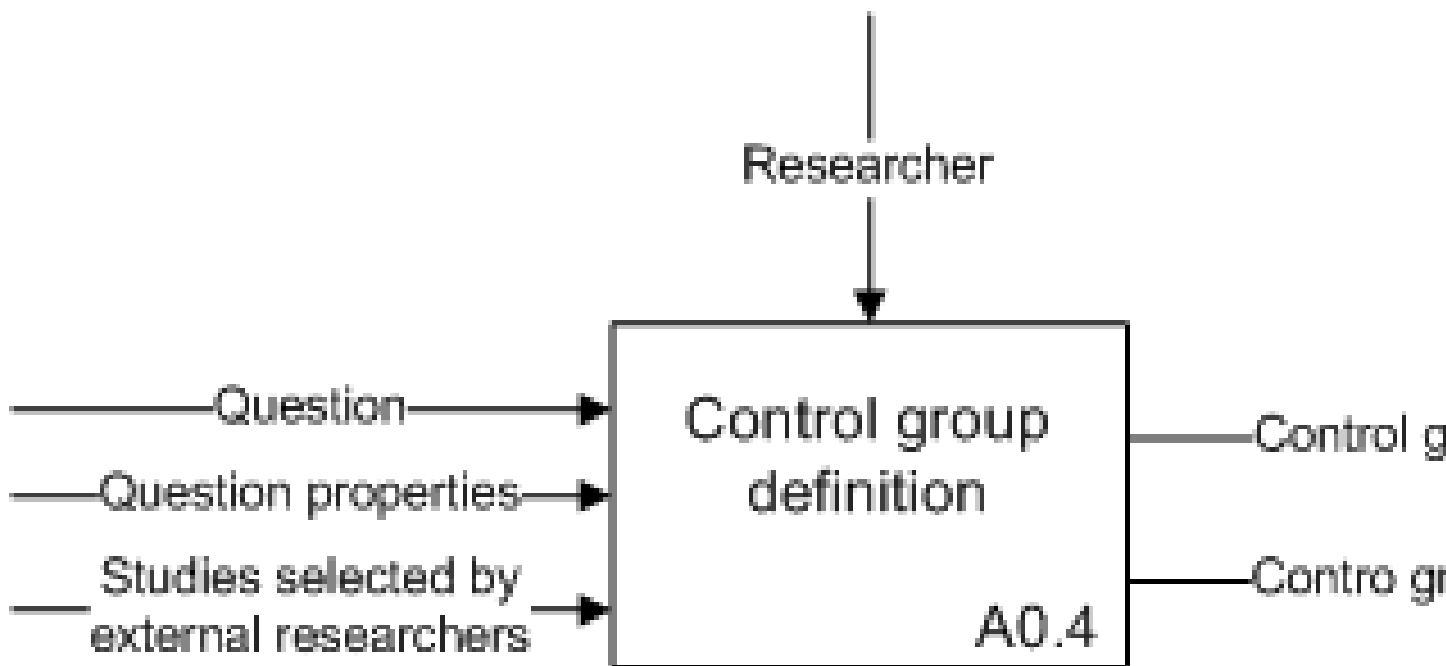


Figura 3.6: Especificação do grupo de controle (*Control group definition*).

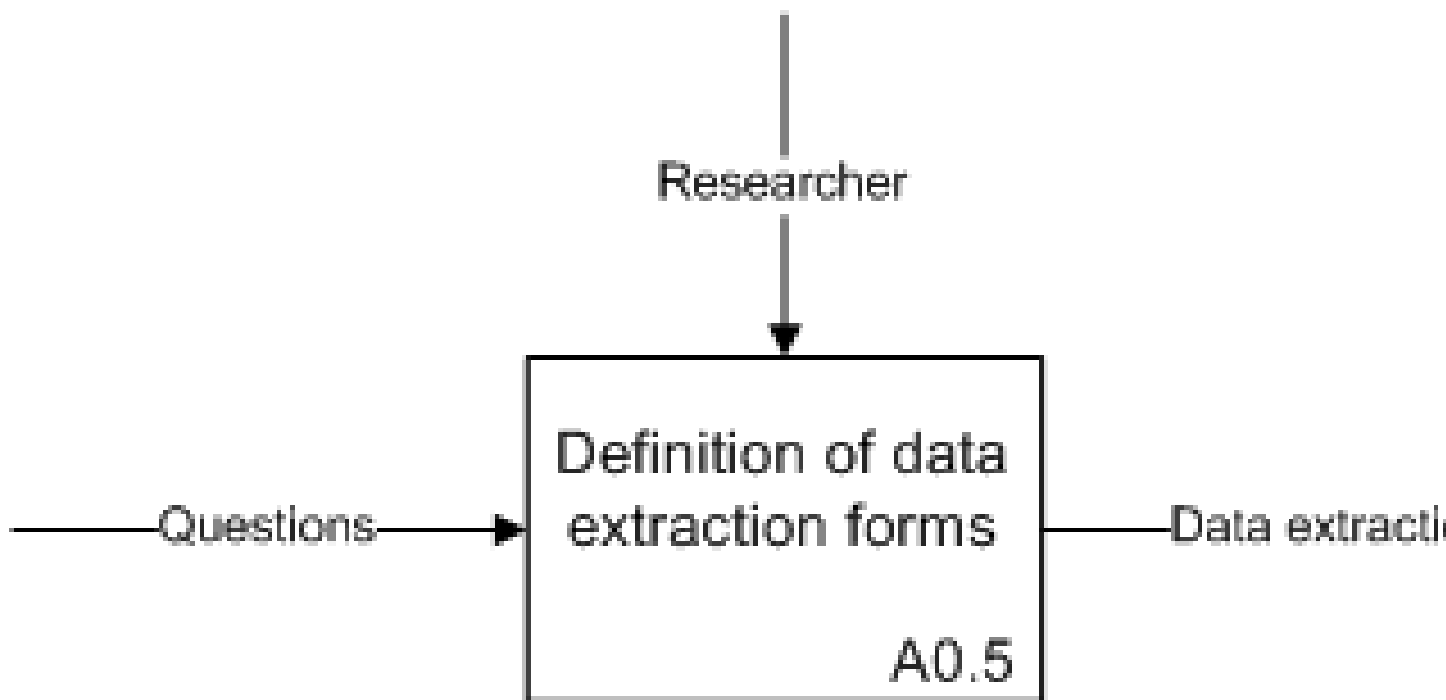


Figura 3.7: Especificação dos formulários para extração de dados (*Definition of data extracton forms*).

as evidências), deve-se reduzir a qualidade do pesquisador. O mesmo vale para os pesquisadores internos. Recomenda-se a seguinte forma para determinar o fator de redução da qualidade do pesquisador:

$$P_s s(r) = z \times \left(\sum_{i=0}^n h_i / ht_i \right), \quad (3.1)$$

$P_s s(r)$ é a penalidade da nota do pesquisador r em função dos fontes selecionadas (*selected sources*) erroneamente, z é o peso a ser atribuído para essa penalidade (a qual fica a cargo do pesquisador escolher),

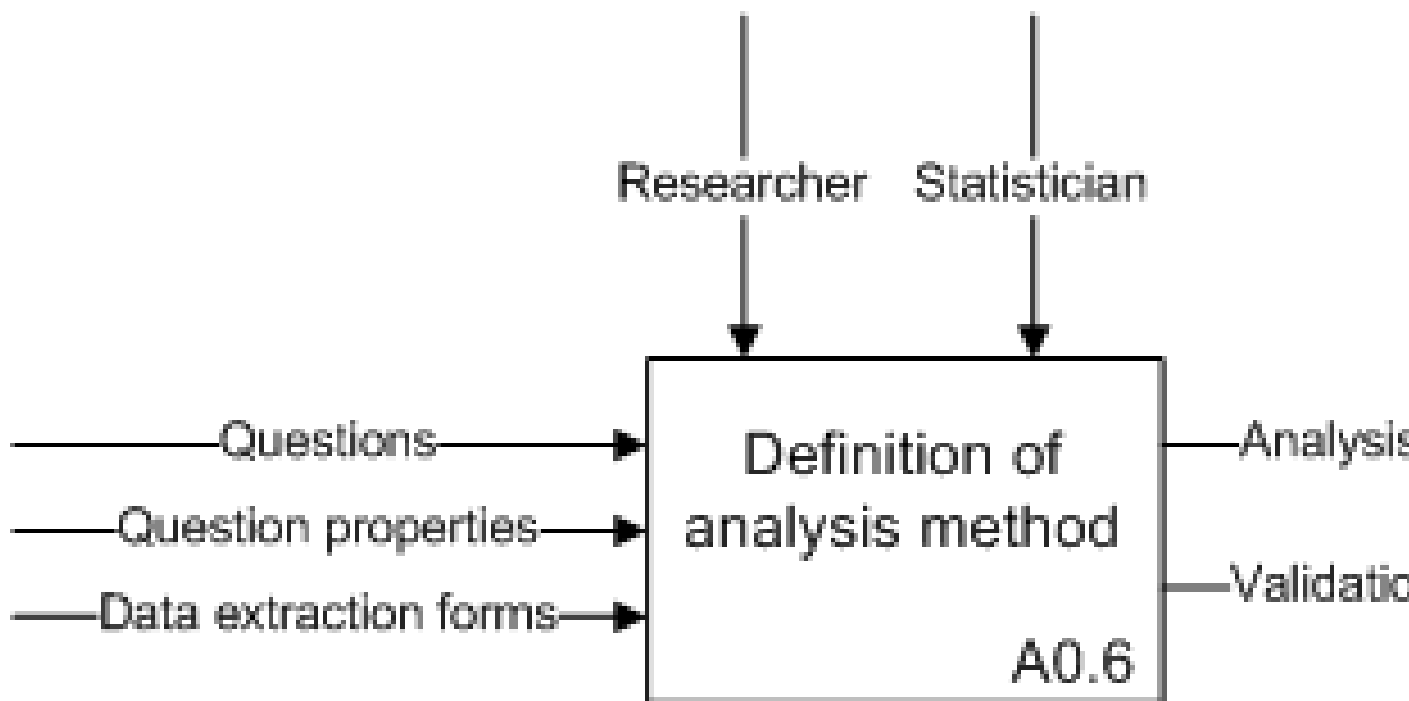


Figura 3.8: Especificação do método de análise e síntese de dados e evidências (*Definition of analysis method*).

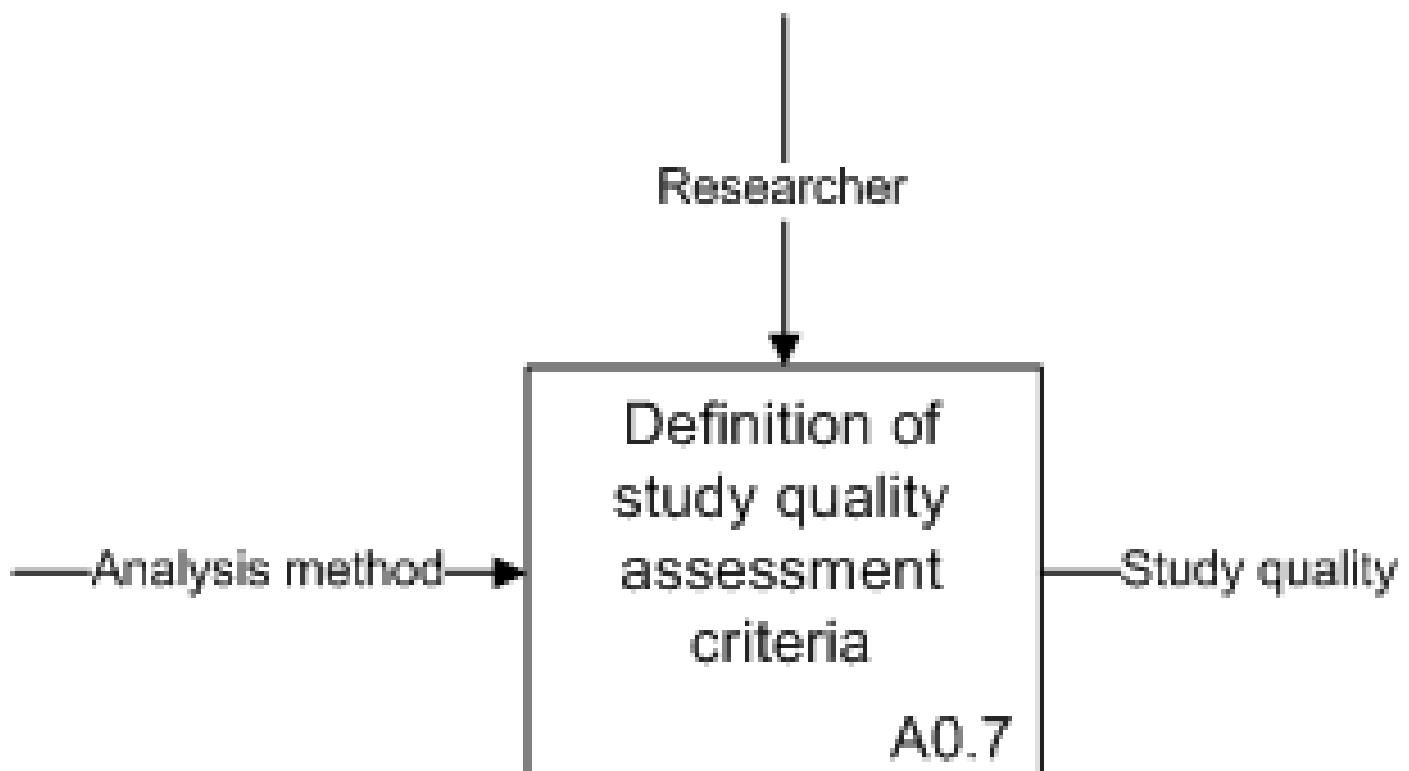


Figura 3.9: Especificação de critérios de avaliação de qualidade de estudos (*Definition of study quality assessment criteria*).

n é a quantidade de fontes sem relevância presentes no grupo de controle, h é quantidade de vezes que o pesquisador r escolheu a fonte i e ht é a quantidade de pesquisadores que selecionaram a fonte i . O fator h_i é acumulativo, ou seja, se, em outras iterações da revisão sistemática, o mesmo pesquisador selecionar a mesma fonte para o grupo de controle, o valor de h_i será não 1, mas 2 (e assim por diante).

A segunda alternativa significa que os critérios de seleção de fonte estão inadequados. Pedidos de alteração dos critérios devem ser criados e registrados em um sistema de controle de configuração.

A busca por novas fontes de estudo deve ser realizada de acordo com os critérios estabelecidos no protocolo. De fato, os critérios não apenas selecionam, mas geram (escolhem) as fontes de estudo. Por exemplo, se um critério determina que a fonte de estudo deva permitir a busca eletrônica de publicações, deve-se escolher um mecanismo de busca que atenda a esse critério. A sugestão das fontes a serem analisadas é oriunda do pesquisador que conduz a revisão sistemática. Geralmente são usadas fontes consagradas da área. No caso de Engenharia de Software, mecanismos de busca como a *ACM Digital Library*, a *IEEE Xplore Digital Library* e o *SciVerse ScienceDirect* são exemplos de fontes utilizadas. Para servir como guia para a escolha das fontes, esses mecanismos de busca devem indexar as fontes presentes no grupo de controle.

Após a seleção de fontes, um conjunto de fontes de estudos primários, que atende aos critérios de seleção (inclusão e exclusão) de fontes, é gerado. Para cada fonte, deve-se estabelecer o procedimento de recuperação de estudos. Geralmente esse processo é conduzido por buscas booleanas simples. Portanto, para cada fonte de estudo, deve-se produzir uma expressão de busca a partir daquela expressão de busca genérica produzida durante a caracterização da questão.

Ao final da seleção das fontes, deve-se avaliar todos os pedidos de alteração de protocolo realizados nesta fase. Para cada pedido, o risco da alteração deve ser avaliado e, com base nele, decidir se a alteração será realizada. Se for decidido pela alteração do protocolo, a revisão sistemática deverá ser reiniciada (alterações de protocolo são possíveis apenas na fase de planejamento).

3.3.3. Seleção de estudos

Os estudos definidos no grupo de controle devem ser analisados com os critérios de seleção estabelecidos no protocolo, de forma análoga à análise feita quanto ao grupo de controle de fontes. A fórmula de avaliação do pesquisador define-se como:

$$P_s s(r) = z \times \left(\sum_{i=0}^n h_i / ht_i \right), \quad (3.2)$$

$P_s s(r)$ é a penalidade da nota do pesquisador r em função dos estudos selecionadas (*selected sources*) erroneamente, z é o peso a ser atribuído para essa penalidade (a qual fica a cargo do pesquisador escolher), n é a quantidade de estudos sem relevância presentes no grupo de controle, h é quantidade de vezes que o pesquisador r escolheu o estudo i e ht é a quantidade de pesquisadores que selecionaram o estudo i . O fator h_i é acumulativo, ou seja, se, em outras iterações da revisão sistemática, o mesmo pesquisador selecionar o mesmo estudo para o grupo de controle, o valor de h_i será não 1, mas 2 (e assim por diante).

O grupo de estudos de controle revisado deve ser agora utilizado para a avaliação das fontes e da adequação do protocolo. Alguns trabalhos apontam que os mecanismos de busca (que são as principais fontes de estudos para as revisões sistemáticas) apresentam restrições quanto à construção das expressões de busca, são inconsistentes quanto aos resultados apresentados e mudam suas características e implementação com frequência. Dessa forma, as regras estabelecidas para a construção de expressões de busca específicas, bem como os procedimentos para a realização das buscas, precisam ser alteradas e validadas para toda revisão sistemática.

Uma forma de avaliar as fontes é verificar as propriedades de recuperação de estudos (precisão e *recall*). Espera-se que todos os estudos do grupo de controle de estudos sejam recuperados das fontes de estudo selecionadas utilizando as expressões de busca específicas. Caso a precisão seja inferior a 100% e a taxa de

recuperação (*recall*) alta (por exemplo, superior a 30%), alterações devem ser realizadas quanto à expressão de busca para aumentar a precisão, reduzir o *recall*. Essas alterações podem ser quanto aos operadores específicos da expressão de busca, alterar os critérios de seleção de fontes ou até mesmo a adição de termos complementares (o que implica em alterar as propriedades da questão de pesquisa).

Definidas as expressões de busca e os procedimentos a serem utilizados para cada fonte, realiza-se a recuperação de estudos. Para cada estudo, deve-se obter, no mínimo, os dados necessários para a aplicação dos critérios de seleção de estudos (referência bibliográfica e resumo).

Para os estudos recuperados, deve-se aplicar os critérios de seleção, conforme definido no protocolo. Ao final da atividade, para todos os estudos selecionados, o texto completo das publicações referentes a cada estudo deve ser obtido (caso contrário a extração dos dados não será possível).

3.3.4. Avaliação de qualidade

De acordo com o método de análise escolhido para a síntese das evidências dos estudos selecionados, é necessário que os estudos possuam uma qualidade mínima. Esta atividade define se a qualidade é suficiente e, caso não o seja, define métodos alternativos de síntese que podem ser utilizados.

3.3.5. Extração de dados

A extração de dados consiste na leitura dos estudos selecionados e a extração dos dados, conforme instruções do protocolo e os formulários de extração de dados.

3.3.6. Análise e síntese de dados

Extraídos os dados, é possível aplicar o método de análise e síntese de evidências estabelecido no protocolo. Ao final da atividade, obtém-se a síntese das evidências, a partir das quais é possível estabelecer a resposta para a questão da qual trata a revisão sistemática.

Cabe também a análise estabelecer as ameaças às validades internas, externas, de conclusão e de construção da revisão sistemática, utilizando, para isso, o protocolo original da iteração corrente da revisão e os pedidos de alteração realizados.

3.3.7. Empacotamento ou documentação

Uma revisão sistemática deve ser documentada de modo a permitir sua replicação e a cuidadosa análise de sua construção e condução. Um elemento essencial na documentação é o protocolo da revisão sistemática. Além disso, pedidos de alteração devem ser documentados e toda decisão tomada quanto a esses pedidos deve ser descrita, com especial atenção quanto a mensuração dos efeitos contrários e favoráveis quanto ao resultado da revisão sistemática.

3.4. Considerações quanto ao processo iterativo

O processo de revisão sistemática pode ser resumido em definição do protocolo, execução e empacotamento (Figura 3.16). Em um modelo simplificado de iteração, as atividades de seleção de fontes, seleção de estudos, extração de dados, análise de qualidade e análise de dados podem ser resumidas em uma grande atividade, a execução do protocolo da revisão sistemática.

No entanto, existem várias oportunidades para verificar a qualidade da revisão sistemática e a necessidade de alterações no protocolo, de modo que, ao invés de se executar a iteração completa, é possível terminá-la prematuramente. Por exemplo, ao detectar um problema com os grupo de estudos de controle e os critérios de seleção, pode-se identificar um erro no protocolo. Assim, ao invés de prosseguir com a revisão, a atual iteração é encerrada e uma nova é iniciada.

- Identificação da necessidade ou não de uma nova iteração.
- Realização de uma nova iteração, reiniciada a partir da etapa de seleção de fontes de estudo.

As iterações também são úteis para realizar incrementos de uma revisão sistemática, ou seja, a realização de uma revisão com o mesmo protocolo, mas com estudos selecionados em um novo período. Embora não faça sentido realizar revisões sistemáticas em curtos períodos, a realização de revisões a cada cinco anos pode demonstrar o quanto as soluções quanto a um problema em específico evoluíram e determinar novas direções de pesquisa. Em particular na área de Engenharia de Software, em que são raros os problemas cuja solução é única, a condução de repetidas revisões sistemáticas é desejável.

No entanto, as iterações devem ocorrer com critérios, com atenção aos pedidos de alteração feitos durante a execução da revisão. De modo a não comprometer a validade da revisão, esta deve ser (1) completamente reiniciada, (2) continuada sem alterações no protocolo ou (3) alterada e continuada, com a adoção das alterações em iterações futuras. Cada uma dessas alternativas possui deficiências, como definido a seguir:

- **Reiniciar revisão com alterações:** O esforço para uma nova realização da revisão sistemática, descartando-se os resultados obtidos na execução anterior (anterior à alteração do protocolo) impõe um custo que talvez signifique o cancelamento da pesquisa. Não obstante, o reinício da revisão não garante a ausência de viés da revisão, dado que as alterações, por serem conseqüência da extração ou síntese dos dados, podem influenciar, positiva ou negativamente, na escolha e análise dos estudos. Por exemplo:
 - A alteração do protocolo devido a uma limitação do mecanismo de busca não é uma alteração tendenciosa (desde que a seleção seja novamente executada).
 - A alteração do protocolo devido a um novo termo encontrado durante a busca, o qual é sinônimo daquele adotado pela comunidade, não é uma alteração tendenciosa (desde que a seleção seja novamente executada).
 - A alteração do protocolo para a indicação de um novo dado a ser extraído é uma alteração potencialmente tendenciosa, dado que os dados a serem extraídos estão relacionados a questão e não deveriam depender do conteúdo dos estudos selecionados e analisados.
 - A alteração do protocolo para a alteração de uma questão secundário é uma alteração tendenciosa, dado que a questão deve ser definida independentemente dos estudos selecionados e analisados.
 - A alteração do protocolo para a alteração da questão primária é uma alteração não apenas tendenciosa, como invalidadora da revisão sistemática em execução.
- **Continuar sem alterações:** Se a resposta à questão primária não for comprometida pelo protocolo, não é necessária alterá-lo. Posteriormente uma nova revisão sistemática pode ser executada, considerando tais alterações.
- **Continuar com alterações:** Conforme a alteração sugerida e o seu possível impacto na revisão sistemática (conforme apresentado no item “Reiniciar revisão com alterações”), é possível aplicar a alteração na revisão em andamento, dado que os riscos da alteração sejam registrados e controlados durante o restante da revisão.

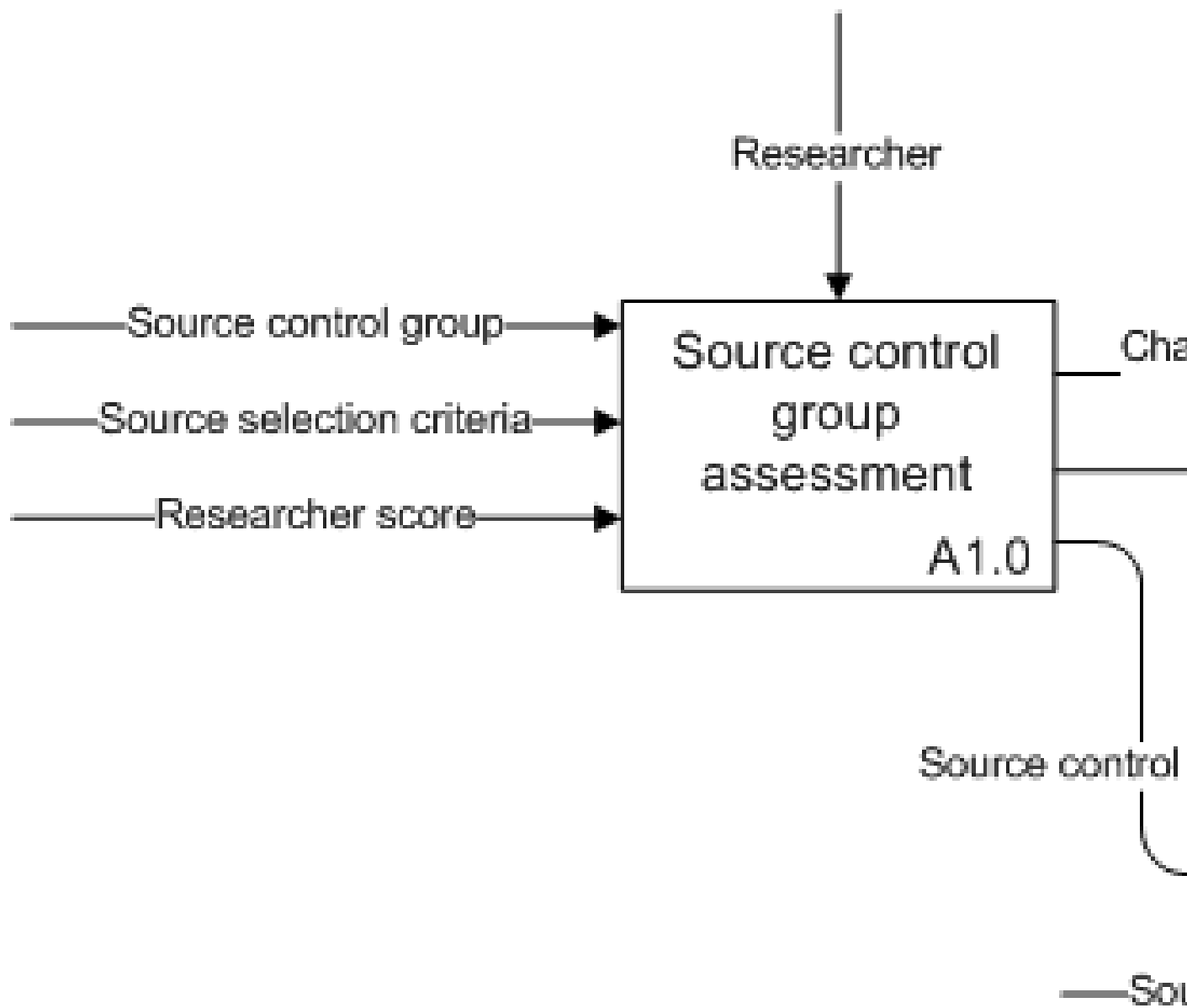
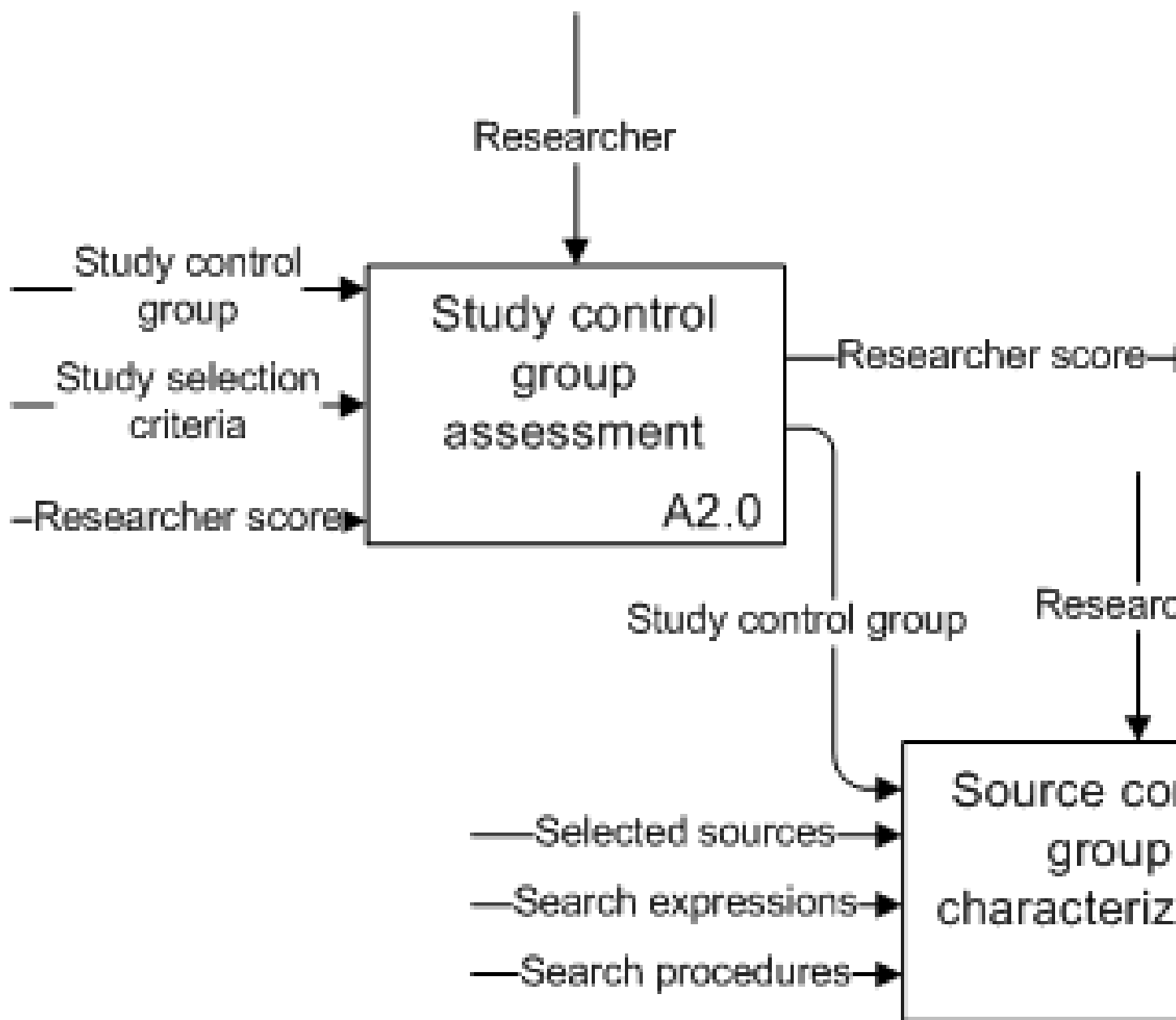


Figura 3.10: Seleção de fontes (*Source selection*).



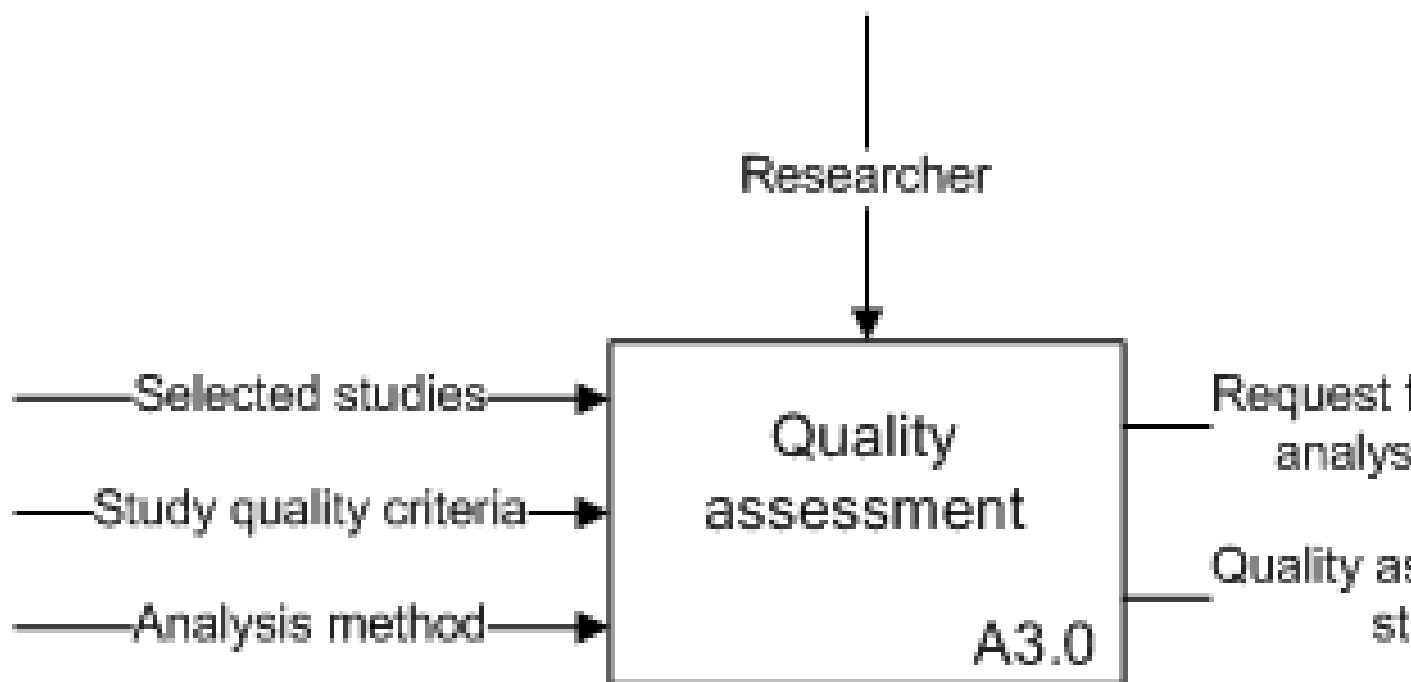


Figura 3.12: Avaliação de qualidade dos estudos selecionados (*Quality assessment*).

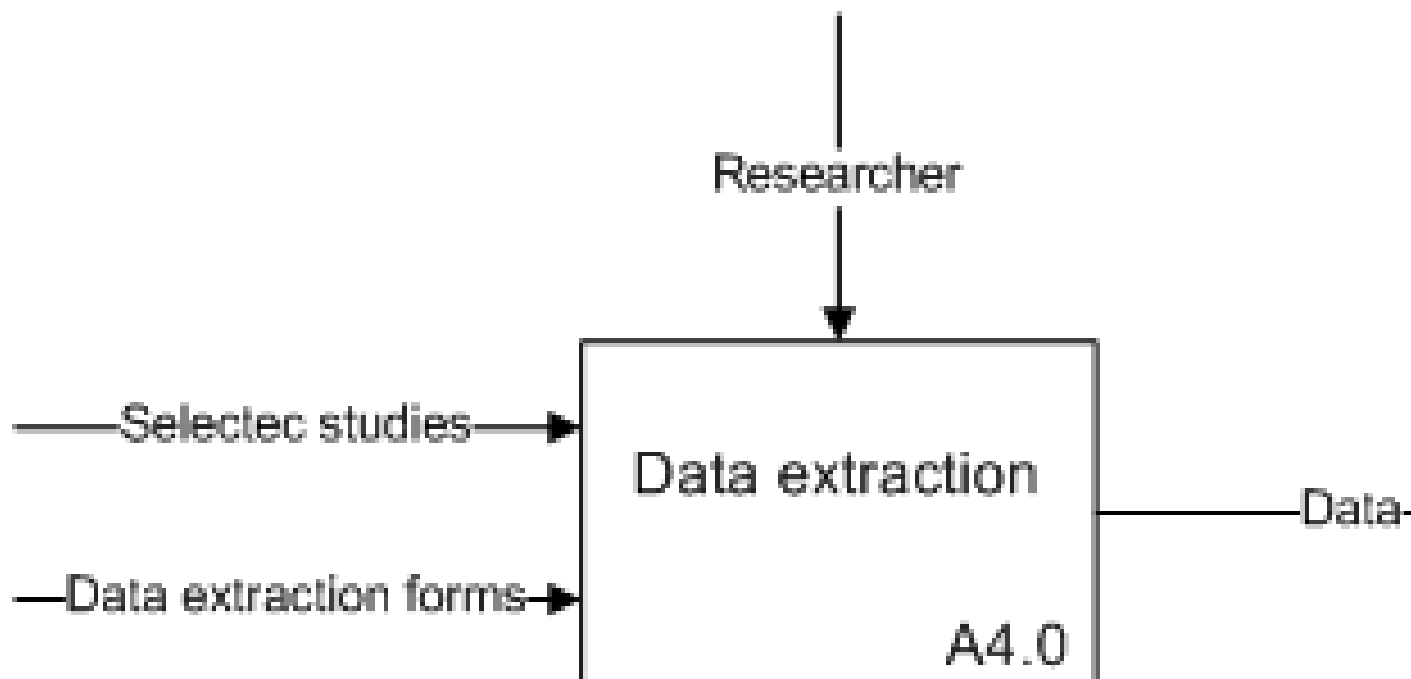


Figura 3.13: Extração de dados dos estudos selecionados (*Data extraction*).

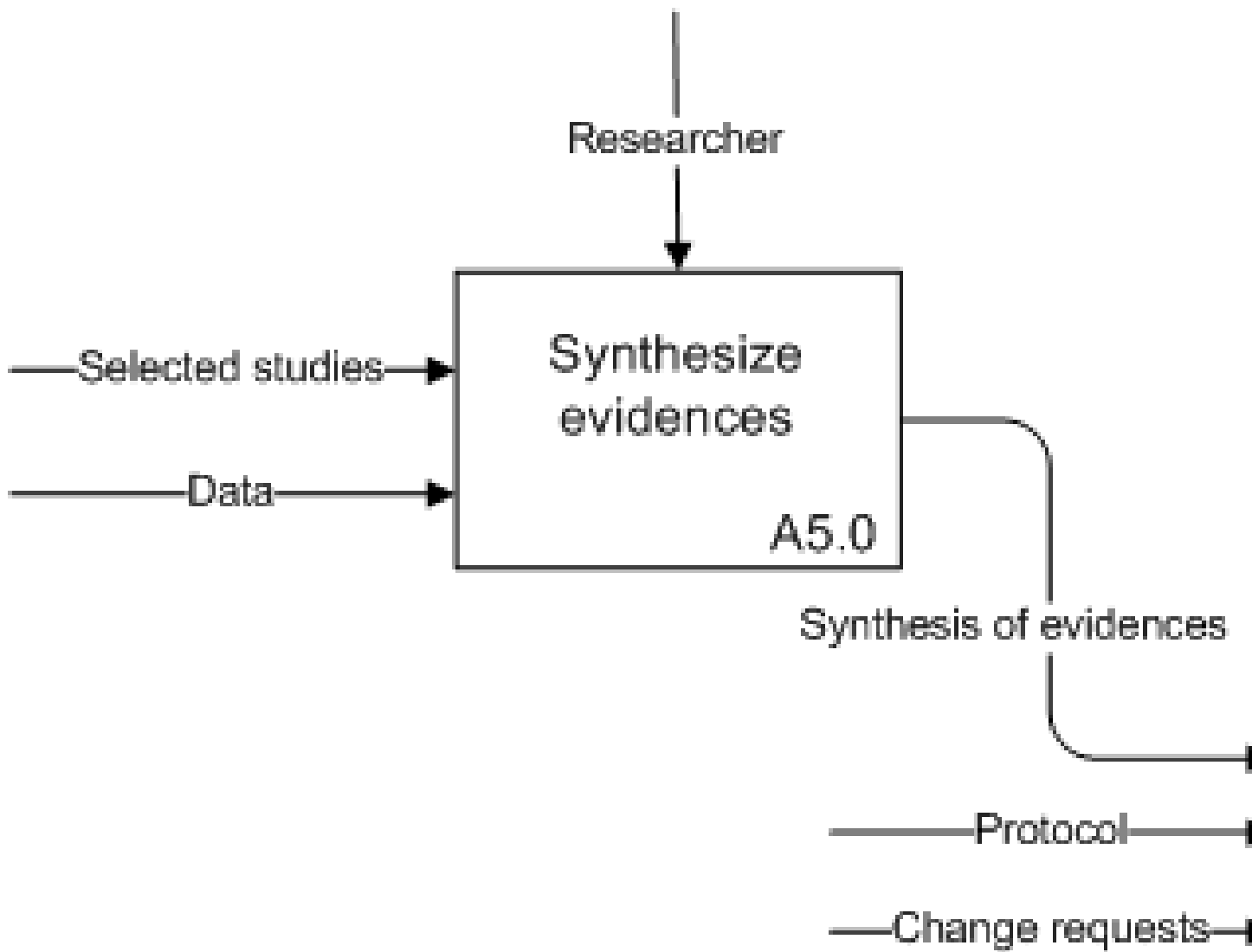


Figura 3.14: Síntese de evidências e análise dos resultados (*Analysis*).

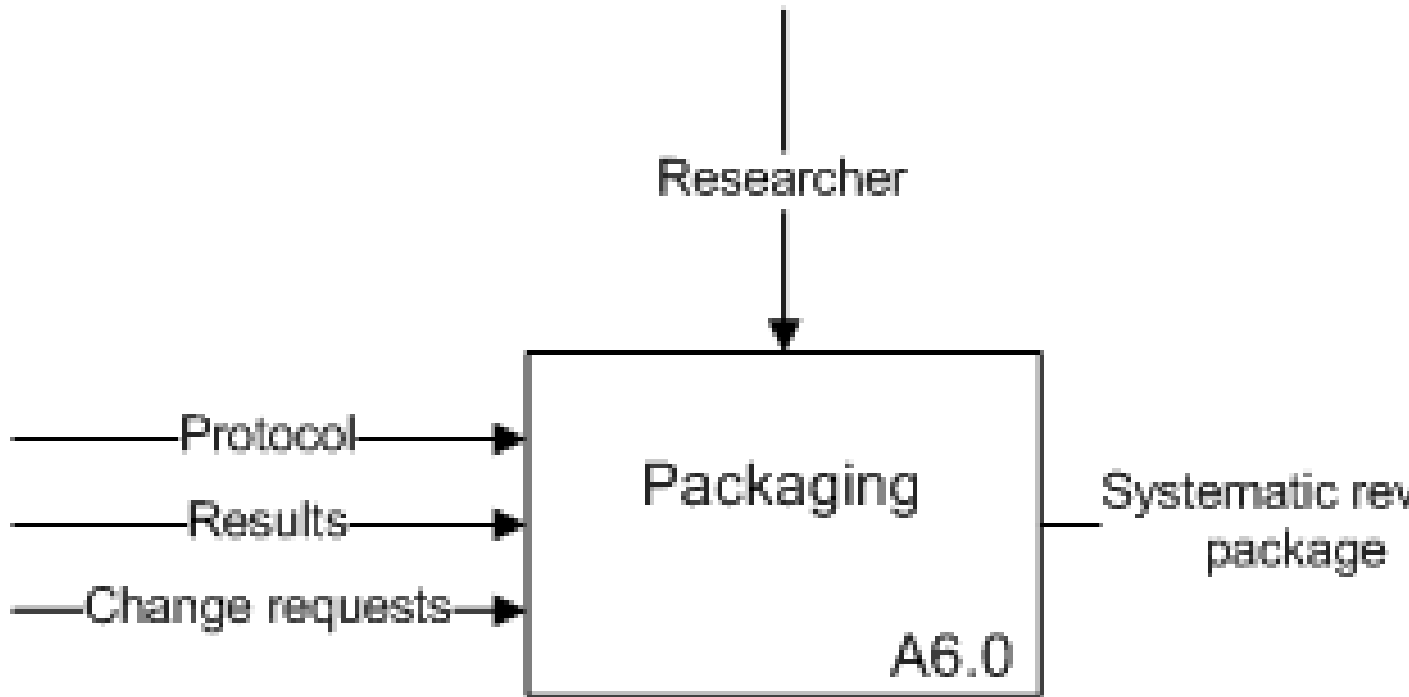


Figura 3.15: Documentação e empacotamento da revisão sistemática (*Packaging*).

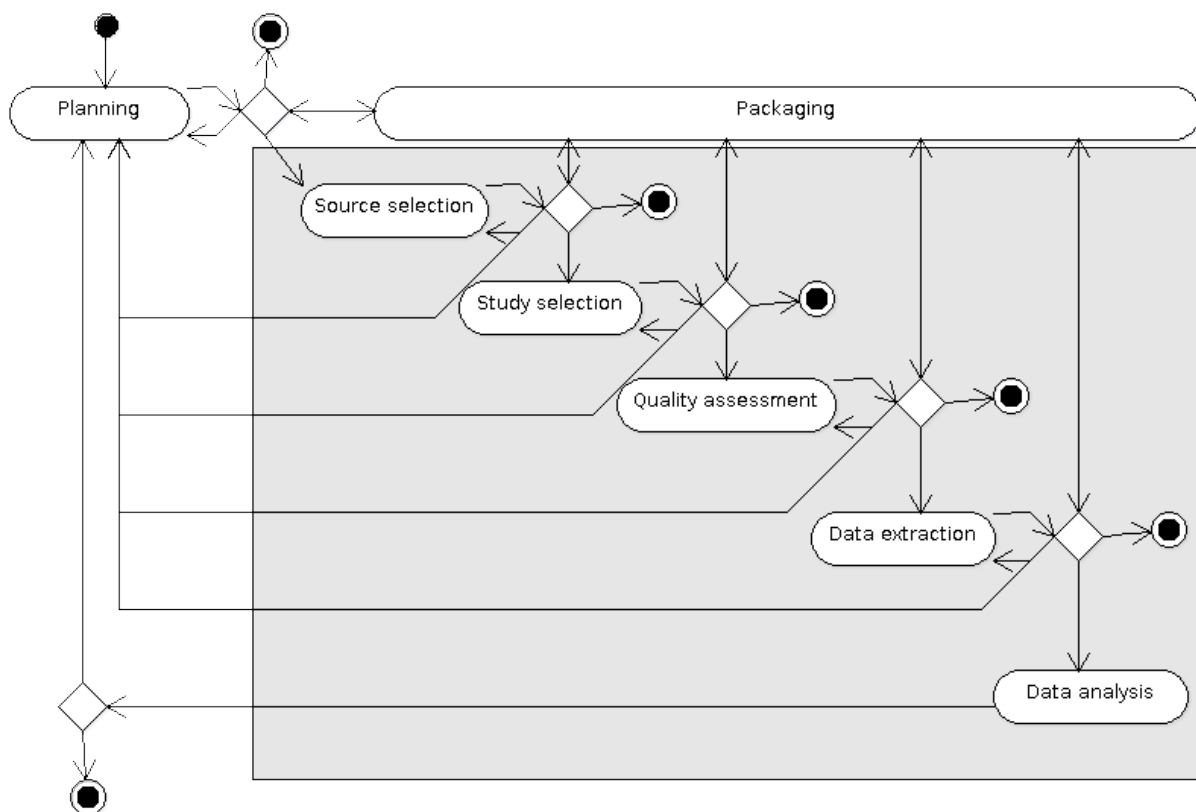


Figura 3.16: Processo de revisão sistemática com agrupamento de diversas fases na macro-fase *Execução*.

Planejamento

No planejamento, define-se o protocolo da revisão sistemática. O objetivo do protocolo é prover uma arcabouço para a execução da revisão sistemática para a qual ele foi originalmente desenvolvido e, principalmente, para a avaliação do rigor e correção e para assegurar a replicabilidade da revisão sistemática (SACKS *et al.*, 1987).

Em outras palavras, o protocolo é mais um mecanismo, dentre outros presentes em estudos experimentais, para o estabelecimento de um grau adequado de qualidade dos resultados (MILLER, 2000). O protocolo de uma revisão sistemática, conforme apresentado na Figura 4.1, organiza-se em:

- **Definição da questão:** Definição da pergunta a ser respondida pela revisão sistemática.
- **Seleção das fontes:** Requisitos para as fontes a serem pesquisada para obtenção dos estudos primários e procedimentos para a seleção.
- **Seleção dos estudos:** Requisitos para a obtenção dos estudos a partir das fontes selecionadas, requisitos e procedimentos para a seleção dos estudos primários.
- **Avaliação de qualidade:** Determinação da qualidade dos estudos primários selecionados.
- **Extração de dados:** Requisitos para a obtenção dos dados dos estudos primários selecionados e formulários para o armazenamento dos dados.
- **Análise e resumo dos resultados:** Análise dos dados extraídos e formas para a apresentação do resultado e da análise de validade da revisão sistemática.

O protocolo, principalmente sua questão, é parte essencial para uma revisão sistemática válida e bem sucedida. As subseções a seguir definem detalhadamente cada uma das partes do protocolo e quais os procedimentos recomendáveis para sua definição.

4.1. Definição da questão

A definição da questão é considerada a parte mais crítica da revisão sistemática (e, de fato, de qualquer trabalho científico). A partir da questão, elaboram-se os termos de busca e os dados a serem extraídos dos

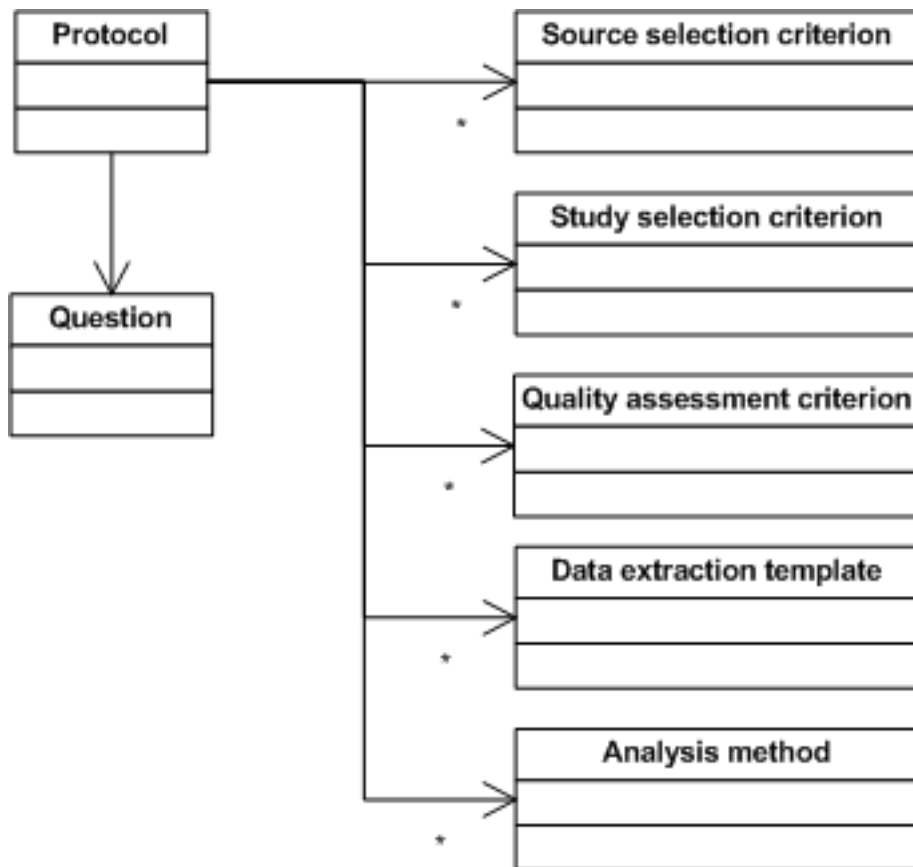


Figura 4.1: Elementos do protocolo de uma revisão sistemática.

estudos.

Uma questão bem formulada deve conter três partes: o elemento sob estudo, a população e os resultados (SACKETT *et al.*, 2000). Kitchenham *et al.* (2004) adaptam essa estrutura para a Engenharia de Software, conforme exposto a seguir:

- **População:** A população pode ser especificada em várias dimensões: tipos de usuários, experiência dos usuários, área de aplicação, tipo de problema.
- **Elemento sob estudo ou intervenção:** Em Engenharia de Software, o elemento sob estudo é a tecnologia de interesse (por exemplo, métodos ágeis, métodos para estimativa de custos, especificações baseadas em contrato).
- **Resultados:** Os resultados são os fatores, dos estudos, considerados importantes (por exemplo, melhora na qualidade do software, redução do custo de produção, etc). Tanto em Medicina quanto em Engenharia de Software, geralmente qualquer tipo de resultado é importante.

Quanto a população, alguns autores optam por não restringí-la excessivamente, dadas as características intrínsecas da disciplina e o risco de perda de evidências relevantes. No entanto, conforme visto em Seção 2.1.2, a população de um estudo experimental em Engenharia de Software (e, portanto, para revisões sistemáticas), envolve um dos quatro elementos: pessoas, produtos, problema ou processo.

Revisões sistemáticas, em tese, sempre observam a população quanto a uma intervenção, ou seja, alguma variação das variável de tratamento. No entanto, algumas revisões sistemáticas (especialmente mapeamento sistemáticos) não objetivam a análise de uma variável específica, mas em observar os elementos da (amostra da) população e caracterizá-los. Por exemplo, na óptica das publicações que descrevem os estudos, pode-se

caracterizar os principais veículos de publicação (conferências e revistas), a forma de publicação (artigos em revista, artigos em conferências, relatórios técnicos), autores, grupos de pesquisa. Sob o ponto de vista do estudo em si, pode-se caracterizar as variáveis definidas em cada estudo primário analisado, identificando os parâmetros e fatores.

No entanto, a sugestão de caracterização da questão da revisão sistemática proposta por Kitchenham *et al.* (2004) não define o controle do estudo experimental e, de fato, não explicita qual a questão a qual a revisão efetivamente trata. Mian *et al.* (2005) sugerem uma estrutura mais detalhada, organizando-a em foco, qualidade e amplitude:

- **Foco:** Define o escopo de interesse da revisão sistemática, ou seja, seus objetivos.
- **Qualidade e amplitude:** Define o contexto ao qual a revisão se aplica e a questão que o estudo deve responder.
 - **Problema:** Breve descrição do contexto da pesquisa.
 - **Questão:** Pergunta a ser respondida pela revisão. Se a questão for muito vaga (ampla), ela deve ser dividida em várias questões secundárias, de modo a melhor delinear o objetivo da pesquisa.
 - **Palavras-chaves e sinônimos:** Termos chaves para encontrar os estudos primários que permitirão a resposta à pergunta proposta pela revisão.
 - **População:** Grupo acometido pela intervenção que será observado.
 - **Intervenção:** O que será observado no contexto da revisão sistemática.
 - **Controle:** Conjunto de dados iniciais que o pesquisador já possui e será utilizado para controle da revisão sistemática.
 - **Efeito:** Tipos de resultados esperados ao final da revisão sistemática.
 - **Medida da saída:** Métricas a serem utilizadas para medir o efeito da intervenção.
 - **Aplicação:** Tipos de pessoas e cargos que se beneficiarão dos resultados da revisão sistemática.
 - **Projeto experimental:** Especificação do método para síntese a ser utilizado (meta-análise estatística, síntese meta-etnográfica, síntese conceitual narrativa).

Esta última organização da questão é mais rigorosa. Alguns itens, como controle, medida de saída e aplicação, não precisam ser definidos, principalmente em se tratando de mapeamentos sistemáticos (que é o caso comum para a caracterização do estado da arte de uma área tecnológica). Em especial, a medida de saída é de difícil caracterização em Engenharia de Software, seja pela quantidade de variáveis, pelo tipo das variáveis e pela falta de padronização das medidas realizadas.

Em comum, todas as três identificam a população, a intervenção e o controle. Em função disso, escolheu-se a técnica Problema ou População, Intervenção, Comparação e Resultados esperados (PICO) para a construção da questão. A aplicação da técnica é simples: define-se o problema ou a população afetada pelo problema. Em medicina, por exemplo, “paciente portador de diabetes tipo II e com feridas no pé do tipo neuropática¹”. Para Engenharia de Software, um exemplo seria “projetos de software livre iniciados nos últimos cinco anos”.

A intervenção é a ação ou observação de interesse realizada no problema ou população. Exemplos: “uso de agente tóxico ou curativo para ferida que contivesse prata em sua composição”, “adoção do modelo de gerência de projeto do ditador-benevolente”.

A comparação (ou controle) é a intervenção padrão, mais utilizada ou a ausência de intervenção a qual deseja-se confrontar a intervenção anteriormente definida (SANTOS *et al.*, 2007). Exemplos: “não utilização de curativos, uso de curativos sem prata em sua composição”, “sem gerência de projeto, modelo tradicional (gerente/funcionário) e modelo ágil”. Caso não seja possível definir um controle dessa forma (por exemplo, em um mapeamento sistemático), deve-se definir as características de um conjunto de estudos

¹ Exemplo médico retirado de Santos *et al.* (2007) (SANTOS *et al.*, 2007 apud BERGIN; WRIGHT, 2006).

que servirá de referência para a caracterização a ser construída pelo mapeamento (assume-se que estudos identificados pela revisão sistemática suportarão ou confrontarão os estudos do grupo de controle, o que cumpre o propósito esperado de uma comparação/intervenção padrão). Por exemplo, para um mapeamento sistemático, o controle pode ser “um conjunto de estudos sobre gerência de projetos de software livre”.

Os resultados esperados (*outcomes*) são exatamente o que o nome diz: os resultados normalmente esperados quanto ao problema ou população com a aplicação da intervenção. Exemplos: “proporção de feridas curadas, mudança na área total da ferida, tempo necessário para cicatrização”, “tempo entre lançamento de novas versões do software, volume de discussão nas listas de *email*, quantidade de alterações no código-fontes”. Observe que sempre são resultados otimistas, porém observáveis e quantificáveis. Não faria sentido indicar “projetos de software bem sucedidos”, dado que não temos como medir essa característica diretamente. Uma revisão sistemática não se propõe a extrair do artigos dados mais amplos e abstratos do que aqueles que estão efetivamente descritos: esse é exatamente o limite do definível em resultados esperados.

Com a definição da PICO, a questão a ser respondida pela revisão sistemática pode ser criada ou interpretada com pouca margem para dúvidas. Por exemplo: “Qual o efeito de curativos e agentes tópicos contendo prata em sua composição no tratamento de feridas de pé em diabéticos?”, “Qual o efeito o modelo de gerência de projeto do ditador-benevolente nos projetos de software livre iniciados nos últimos cinco anos?”.

Quando documentada a questão e os seus atributos, deve-se descrever, brevemente, o problema a ser analisado e que soluções se buscam. Disto deve-se identificar, claramente, a justificativa para a realização de uma revisão sistemática para tratar do problema, procedendo-se então à definição da questão e demais elementos, focando-se assim a pesquisa.

4.2. Definição do controle

A comparação (ou controle) é a intervenção padrão, mais utilizada ou a ausência de intervenção a qual deseja-se confrontar a intervenção anteriormente definida (SANTOS *et al.*, 2007). Geralmente ela consiste em:

1. Elementos da amostra da população que não sofreram a mesma intervenção que os demais elementos da amostra.
2. Perguntas cruzadas, ou seja, que servem para validar o questionário e identificar tendências introduzidas no projeto de um questionário.
3. Grupo de estudos que representam (ou deveriam representar) o *status quo* da área de pesquisa.

O grupo de controle varia de acordo com o tipo de estudo encontrado na revisão. Por exemplo, para experimentos e até mesmo estudos de casos, uma comparação como a do Item 1. pode ser utilizada. Para levantamentos (*surveys*, algo na linha do Item 2. é mais adequado. Finalmente, para mapeamentos sistemáticos, o Item 3. é a escolha mais usual.

Infelizmente, a quantidade de estudos primários é insuficiente para a quantidade de variáveis envolvidas e resultados esperados e estudados em Engenharia de Software. Seria possível associar vários estudos quanto ao sucesso ou não de um projeto de software, mas a relação causal entre uma única ou poucas intervenções (fatores) e o efeito (sucesso de software) é muito tênue, para não dizer impossível de ser estabelecida. Por exemplo, atribuir o sucesso de um projeto de software em função do uso de programação em pares seria incorreto, mesmo que vários estudos apontem que projetos que utilizam programação em pares seja bem sucedidos: muitas outras variáveis estão presentes e influenciam o resultado final. A identificação de todos os parâmetros (contexto) em pesquisas de Engenharia de Software também é inviável, dada a forte influência humana (e não apenas técnica) no resultado final. Ou seja, a definição do controle para estudos primários, principalmente para estudos de caso e experimentos, é inviável (dado que ele restringiria muito o foco da

busca de evidências, correndo o risco de não encontrar publicações relevantes). De fato, um dos objetivos da Engenharia de Software Experimental é suprir essa deficiência nas pesquisas, incentivando o uso de estudos empíricos.

A definição de um grupo de controle para mapeamentos sistemáticos é mais simples, dado que, por consequência natural das pesquisas, o pesquisador possui um corpo de estudos que servem como base (ou motivação) para o mapeamento sistemático. No entanto, deve-se precaver da utilização, como controle, dos estudos utilizados para formular a questão da revisão sistemática. A utilização dos mesmo estudos como justificativa/motivação e para avaliar a intervenção e a hipótese assumida na revisão sistemática insere um viés: a tendência dos resultados obtidos estar de acordo com o grupo de controle e, principalmente, com o sucesso da intervenção proposta. Por exemplo, caso o objetivo seja caracterizar o estado da arte de uma área de pesquisa, os estudos de controle poderão limitar ou induzir os resultados àqueles por eles evidenciados. Talvez seja realmente a questão de que o grupo de controle represente fielmente o estado da arte, mas não poderá ser eliminada a hipótese do resultado ter sido alcançado pela não inclusão ou detecção de estudos relevantes, porém contraditórios em relação ao grupo de controle (lembrando que o grupo de controle influencia a seleção de fontes e de estudos). Dessa forma, os estudos incluídos no grupo de controle não devem ser os mesmos utilizados para formular a questão de pesquisa ou o grupo de controle deve ser fornecido por pesquisadores independentes. Finalmente, na avaliação da validade dos resultados, deve-se aferir a qualidade e a influência dos resultados obtidos: quanto maior a predominância de evidências dos estudos de controle, menor é a força e a qualidade dos resultados obtidos.

O controle de revisões sistemáticas pode-se realizar por listas de estudos ou de fontes de estudos relevantes. Idealmente, ambas devem ser utilizadas. Caso uma lista de estudos tenha sido definida, devem ser incluídas fontes que indexem os estudos do controle. Especificamente para a área de Engenharia de Software, algumas fontes de estudo são tradicionalmente incluídas:

- IEEE: <http://ieeexplore.ieee.org>
- ACM: <http://www.acm.org/dl>
- Scirus (Elsevier): <http://www.scirus.com/>
- SpringerLink: <http://www.springerlink.com/>
- ISI Web of Knowledge: <http://portal.isiknowledge.com>

Recomenda-se a utilização de, no mínimo, o conjunto acima de fontes como controle inicial. Posteriormente, após a aplicação dos critérios de seleção para fontes, alguma dessas fontes podem ser excluídas, obviamente.

4.3. Definição dos termos (palavras-chaves)

A definição das palavras-chaves a serem utilizadas para a busca de estudos nas fontes selecionadas geralmente é baseada nos atributos da questão, os quais são estabelecidos pelo PICO. Para cada atributo, as principais palavras (tópicos), geralmente substantivos, são extraídos.

Algumas áreas, como a medicina, possuem um vocabulário estabelecido (e respeitado) para cada termo definido como atributos da questão (PICO). Aos termos pertencentes a esse vocabulário utiliza-se o nome de “termos controlados”. Os demais termos não pertencentes a esse vocabulário (palavras, sinônimos, variações de grafia, siglas) são classificados como “termos não-controlados”. O ideal é a utilização dos termos controlados sempre que possível.

Infelizmente, em computação não existe essa padronização na nomenclatura (apesar da existência de dicionários como o IEEE 610.12 (IEEE, 1990)). Esforços estão sendo direcionados no estabelecimento de

ontologias para as várias áreas da computação, mas esses trabalhos ainda são incipientes. Logo, é necessária a adoção de estratégias simples para a obtenção das palavras-chaves (por exemplo, utilizar os substantivos das questões). Por exemplo, considere os atributos abaixo:

- **População:** Projetos de software livre.
- **Intervenção:** Adoção do modelo de gerência de projeto ditador-benevolente.
- **Controle:** Modelo tradicional de gerência (ditador tradicional).
- **Resultados:** Redução no tempo entre lançamento de novas versões de software.

Como se trata de uma revisão sistemática no âmbito de Engenharia de Software, sem termos controlados, os termos devem ser obtidos diretamente da questão. Nesse caso, tomam-se os principais substantivos como termos:

- **População:** projeto, software livre.
- **Intervenção:** modelo de gerência, ditador-benevolente.
- **Controle:** modelo gerência, tradicional, ditador.
- **Resultados:** lançamento de versão.

Para o caso de termos controlados, não é necessário procurar palavras semelhantes. No entanto, para termos não-controlados, deve-se complementar as palavras-chaves com termos semelhantes. Segue uma estratégia para a determinação dessas palavras:

- Os termos devem ser traduzidos para os idiomas dos estudos primários a serem investigados.
- Para cada termo, deve-se identificar sinônimos, nos seu respectivo idioma, na literatura da área, como, por exemplo:
 - Livros: Aqueles adotados em cursos de graduação e pós-graduação.
 - Artigos publicados em periódicos e conferências de impacto elevado.
 - Vocabulário da área: Algumas organizações definem glossários ou o vocabulário da área. Por exemplo, a IEEE possui a norma IEEE 610.12 para a terminologia em Engenharia de Software.
 - Consultas a especialistas.

Considerando-se o exemplo anterior e a estratégia então definida, os seguintes sinônimos foram agregados aos termos identificados:

- **População:** projeto, software livre, software de código-aberto
- **Intervenção:** modelo de gerência, ditador-benevolente, déspota esclarecido.
- **Controle:** modelo gerência, modelo tradicional, ditador.
- **Resultados:** lançamento de versão.

Esses termos serão posteriormente utilizados para a busca de estudos, mais especificadamente para a construção de expressões de busca a serem utilizadas nas fontes que são indexadoras de artigos.

4.4. Construção de expressão genérica de busca

Os termos previamente identificados (a partir dos atributos da questão) devem ser combinados com operadores booleanos da seguinte forma:

1. Para cada atributo, os termos a ele associados devem ser combinados com o operador **OR**. Salienta-se

que os termos compostos por mais de uma palavra precisam ser definidos entre aspas, de modo a evitar ambiguidades (evitar que sejam confundidos com termos distintos).

2. Todos os conjunto de termos devem ser combinados com o operador AND

Por exemplo, a partir dos termos definidos na ??, a seguinte expressão genérica de busca é gerada:

```
projeto OR "software livre" OR "software de código-aberto"  
AND  
"modelo de gerência" OR "ditador-benevolente" OR "déspota esclarecido"  
AND  
"modelo gerência" OR "modelo tradicional" OR ditador.  
AND  
"lançamento de versão"
```

Essa expressão de busca genérica será posteriormente especializada para cada fonte identificada. Isso é necessário porque os diferentes mecanismos de busca possuem especificidades quanto à sintaxe das expressões, bem como limitações quanto ao tamanho da expressão, uso de conjunções e disjunções, lematização de palavras, etc.

4.5. Definição dos critérios de seleção das fontes

Nesta atividade definem-se os critérios para a seleção das fontes (mecanismos de busca, bibliotecas, etc) das quais serão obtidos os estudos a serem sintetizados para responder a pergunta estabelecida pela revisão sistemática. Em resumo, os seguintes itens devem ser analisados para a definição de critérios para a seleção de fontes de estudos:

- Facilidade de acesso à fonte.
- Relevância da fonte para a área de pesquisa.
- Mecanismos para recuperação de estudos.
- Idioma das publicações disponíveis.
- Período indexado pela fonte.
- Tipo de publicações indexadas pela fonte.
- Tipo de fonte.

4.5.1. Acesso

Uma característica desejável em uma revisão sistemática é a capacidade de ser replicável. Um requisito para isso é que as fontes dos estudos sejam acessíveis pela comunidade científica. Hoje em dia, isso traduz-se em fontes que tenham presença na Internet.

Outros critérios a serem estabelecidos são o custo de acesso à fonte (preferencialmente gratuito), a facilidade de acesso aos estudos (disponibilidade parcial ou total, em meio digital ou físico).

4.5.2. Relevância

As fontes de estudos primários da população sob investigação devem ter relevância acadêmica. Uma forma de observar isto é pelo fator de impacto das fontes de estudos (periódicos e conferências) ou dos estudos indexados pelas fontes. O fator de impacto das publicações pode ser obtido de sites especializados, como, por exemplo, o CiteSeer (<http://citeseer.ist.psu.edu/impact.html>). Também é possível aferir a qualidade por bases nacionais, tal como o ERA/ARC/Austrália (<http://www.arc.gov.au/era/>) e Qualis/CAPES/Brasil

(<http://qualis.capes.gov.br/webqualis/>). Para os casos em que um fator de impacto não existe, deve-se verificar o rigor da seleção de trabalhos da publicação (revisões por pares, relação entre artigos submetidos e aceitos, etc) e as organizações que apoiam o evento (por exemplo, eventos apoiados pela IEEE e ACM provavelmente são de boa qualidade).

4.5.3. Mecanismos para recuperação de estudos

Um outro requisito importante são as capacidades de busca que a fonte oferece. Espera-se, no mínimo, a capacidade de busca por palavras-chaves. Suportar expressões booleanas também é desejável.

No entanto, algumas fontes, tal como bases de relatórios técnicos de projetos governamentais, relevantes em alguns cenários de pesquisa (GLASS, 1991), não suportam tais buscas e, mesmo assim, devem ser investigadas para a obtenção de resultados adequados. Para definir as fontes não-indexadas que devem ser consideradas, deve-se consultar especialistas do assunto (pesquisadores externos).

4.5.4. Idioma dos estudos

A determinação do idioma dos estudos a serem selecionados é essencial e, talvez, uma das mais importantes barreiras encontradas. Estudos relevantes publicados em idiomas desconhecidos dos pesquisadores são de difícil ou impossível análise, o que acarreta no descarte de evidências (mesmo que elas sejam relevantes). As fontes habitualmente utilizadas em Engenharia de Software indexam estudos em inglês, o que torna esse idioma um forte candidato à seleção. Idiomas nativos também são desejáveis. No entanto, para alguns idiomas (por exemplo, o português), o acesso aos estudos e a ausência de mecanismos indexadores adequados inviabiliza o uso de fontes deste idioma (com exceção das fontes utilizadas como controle, sugeridas pelos pesquisadores).

4.5.5. Período indexado

A revisão sistemática deve recuperar as melhores evidências sobre a questão disponibilizados até o presente momento. Dessa forma, deve-se assegurar que a fonte de estudos indexa trabalhos de períodos contínuos e seja atualizada com frequência. Caso contrário, não é possível afirmar que a revisão contemplou todos os estudos relevantes e recolheu as melhores evidências disponíveis.

4.5.6. Tipos de publicações

Os estudos são documentos por publicações: artigos publicados em revistas ou apresentados em eventos, relatórios técnicos, dissertações, etc. Artigos em Engenharia de Software (e em muitas áreas científicas) geralmente apresentam apenas os aspectos positivos da aplicação de uma técnica, omitindo aspectos negativos e dificuldades encontradas. Para algumas revisões sistemáticas, esse é exatamente o tipo de informação almejado. Logo, o uso de artigos insere um viés na revisão, “desprezando” evidências contraditórias, porém corretas. Felizmente, relatórios técnicos e dissertações, dada a liberdade de redação, podem conter tais informações. A definição de critérios de seleção quanto ao tipo de publicação permite explicitar a necessidade por um tipo específico de publicação.

4.5.7. Tipo de fontes

O principal tipo de fonte utilizado são os mecanismos de busca, mais especificamente as bases indexadas de publicações científicas. Porém, existem outros tipos de fontes, tal como, por exemplo, pesquisadores e as próprias publicações.

4.5.7.1. Mecanismos de busca

São o tipo de fonte mais utilizado. Provê mecanismos de busca, com suporte de diferentes tipos de expressões de busca.

4.5.7.2. Pesquisadores

4.5.7.3. Publicações

Por exemplo, se for considerado um processo de revisão sistemático iterativo, é possível incorporar como fontes de estudos os próprios estudos encontrados em uma etapa anterior. Se um determinado estudo é frequentemente citado por estudos selecionados e analisados, mas estudo esse que não foi encontrado pela busca nas fontes selecionadas na iteração anterior, o próprio estudo poderia ser incluído como um fonte (com a recuperação dos estudos por ele citado e a aplicação de critérios de seleção de estudos).

Pode-ser argumentar que o uso de citações para inclusão introduz um viés, dado que estudos que não foram encontrados pela expressão de busca foram incluídos. No entanto, cabe ressaltar que os estudos recuperados por esse mecanismo serão submetidos aos mesmos critérios de seleção dos demais estudos.

Uma outra alternativa seria a identificação de uma fonte de estudo que contenha o estudo em questão. Assim não existiria o viés da inclusão direta do estudo, permitindo a inclusão de estudos relacionados que pertençam a mesma base. Caso a fonte já esteja incluída, deve-se revisar a expressão de busca específica, a expressão de busca genérica e as palavras-chaves para identificar a razão do estudo não ter sido encontrado/recuperado.

4.6. Definição dos critérios de seleção de estudos

Os estudos encontrados nas fontes selecionados devem ser analisados para utilização efetiva na revisão sistemática. Essa análise dá-se a partir da aplicação de critérios de inclusão e exclusão, definidos de acordo com os propósitos da questão. Esse alinhamento com a amplitude da questão de pesquisa com os critérios permite a redução da heterogeneidade do conjunto de estudos sob análise (MILLER, 2000) (o que contribui para a posterior aplicação de técnicas síntese, tal como a meta-análise estatística).

Os critérios de inclusão e de exclusão especificam os tipos de estudos experimentais, populações e resultados que serão incluídos na revisão (DYBA *et al.*, 2007). Em Engenharia de Software, não existem repositórios que descrevam os estudos de acordo com tais elementos. Embora seja possível, durante a busca de publicações nas fontes de estudo, a definição de requisitos mínimos esperados para uma publicação (por exemplo, que ela seja um artigo completo, escrita em inglês, publicada em uma conferência ou revista arbitrada), esses elementos não se confundem com aqueles especificados com os principais itens abordados pelos critérios. De fato, pode-se argumentar que a exclusão “automática” de publicações com base no tipo de publicação, sua extensão em páginas ou quantidade de palavras ou idioma utilizado (dentro inúmeras outras características) introduz um viés na revisão, dado que: (1) estudos relevantes, que atendam os critérios de

inclusão quanto a estudos experimentais, populações e resultados, podem ser desconsiderados; a indexação dos estudos e os mecanismos de busca não são perfeito e podem classificar publicações erroneamente. No entanto, também é razoável ponderar que: (1) o universo de estudos pode ser superior àquele que é viável avaliar pelo grupo de pesquisadores que conduzem o estudo; e (2) os estudos que porventura sejam encontrados representem uma diferença significativa nos resultados finais.

Expostas essas questões sobre os critérios de inclusão e exclusão, recomenda-se a seguinte estratégia:

1. Identificar estudos sem impor restrições no mecanismo de busca (quanto à quantidade de páginas, tipo de estudo, etc) e comparar com a quantidade de estudos identificados com a imposição de restrições.
 - Se a diferença for inexistente, não utilizar tais restrições nos critérios de exclusão.
 - Se a diferença for reduzida (inferior a 5% ou, no máximo, 10% do total encontrado com restrições), utilizar a versão com restrições (caso a perda de confiabilidade do resultado final seja aceitável) ou utilizar a versão sem restrições.
 - Caso contrário, utilizar a versão com restrições.

Para o caso dos critérios de exclusão serem aplicados por intermédio de restrições nos mecanismos de busca, é recomendável a adoção de medidas que garantam que os estudos excluídos por tal critério sejam revisados:

1. Selecionar, aleatoriamente, dez estudos do conjunto de estudos excluídos.
2. Avaliar os estudos de acordo com os critérios de inclusão e exclusão.
 - Caso algum estudo seja incluído, avaliar os atributos de qualidade do estudo.
 - Caso a soma dos atributos de qualidade dos estudos selecionados seja superior a 1, reconsiderar a exclusão dos estudos pelo critério anteriormente adotado. De fato, sugere-se a adoção de critérios mais brandos, que permitam a inclusão de tais estudos.
 - Caso contrário, considerar satisfatório o conjunto de artigos obtidos originalmente.

Em um primeiro instante, a abordagem então descrita parece introduzir um viés, permitindo a inclusão de mais estudos que o protocolo originalmente determinava. No entanto, uma importante distinção é necessária: os critérios de seleção e exclusão não são alterados durante o processo. De fato, a única diferença é quanto ao momento de aplicação de um subconjunto desses critérios: na identificação e recuperação dos estudos ou na etapa de aplicação dos critérios pelos condutores da revisão sistemática. Caso os mecanismos de busca fossem confiáveis e os estudos fossem descritos de forma que seus atributos diretamente relacionados a experimentos fossem explicitamente definidos, não seria necessária designar essa parte da aplicação de um subconjunto dos critérios aos aplicadores do experimento. No entanto, é de conhecimento que os sistemas de busca possuem limitações (FERRARI; MALDONADO, 2008) e os estudos não são descritos de forma apropriada (BUDGEN *et al.*, 2008; KITCHENHAM *et al.*, 2008), o que torna tais cuidados necessários.

Retomando a questão da definição dos critérios de inclusão e exclusão (ou seja, ignorando-se agora a questão dos mecanismos utilizados para aplicação desses critérios), é necessário definir como eles serão especificados. Neste momento, é importante destacar que os critérios devem aumentar a homogeneidade dos estudos a serem sintetizados. Embora seja desejável que eles sejam rígidos, com o objetivo de reduzir a quantidade de estudos que serão completamente analisados, eles também não devem excluir estudos importantes. Ao mesmo tempo, os critérios não devem ser frágeis a ponto de incluir estudos com hipóteses distintas ou cujos resultados estejam relacionados a variáveis muito distintas entre si².

² Caso o objetivo seja realmente realizar o mapeamento de uma área, é possível abrandar os critérios, neste caso a revisão sistemática é um mapeamento sistemático.

Os critérios refletem as características (amplitude) da questão de pesquisa: tipos de estudos experimentais, populações, resultados. Alguns exemplos de critérios:

- **Critérios de inclusão:**

- Experimentos sobre o desenvolvimento de software utilizando programação em pares.
- Estudos de casos executados na indústria de sistemas embarcados.
- Estudos de caso sobre uso de ferramentas de teste que envolvam alunos de graduação em Computação nos anos de 2000 a 2005.
- Experimentos aleatorizado e cego sobre a efetividade de técnicas para medição no contexto de gerenciamento de software (por exemplo, os gerentes de um grupo de empresas regem o desenvolvimento de seus projetos utilizando medidas fornecidas por um serviço Web, o qual fornece medidas baseadas em uma técnica X ou Y, da qual o gerente não tem conhecimento).

- **Critérios de exclusão:**

- Estudos não experimentais.
- Experimentos observacionais.
- Estudos experimentais cujos resultados não sejam do tipo quantitativo.
- Estudos de caso que envolvam menos de 23 projetos de software ou com mais de 23 projetos, porém não escolhidos aleatoriamente.

A aplicação dos critérios de seleção exige um esforço considerável por parte dos pesquisadores. Uma forma, em estudo, para reduzir o custo da seleção, principalmente para buscas que recuperam muitos estudos (mais de 500), é a utilização de técnicas de mineração visual de texto (VTM – *Visual Text Mining*).

Um possível critério de seleção baseado em *Visual Text Mining* (VTM) é o uso de citação nos artigos para a criação de mapas de documentos. Por exemplo, um artigo, encontrado durante a busca, e que não é citado por nenhum outro artigo, provavelmente não é relevante e, portanto, pode ser excluído. O mesmo pode ser aplicado para a inclusão: se um artigo possui muitas citações para artigos que foram selecionados de acordo com outros critérios de seleção, ele também deveria ser selecionado.

Definidos os critérios de seleção, procede-se à especificação dos procedimentos para a seleção dos estudos também deve ser especificado. Esse procedimento indica como o método de pesquisa (no caso, busca de estudos por palavras-chaves) deve ser executado e como os critérios de inclusão e exclusão devem ser aplicados aos estudos encontrados.

A aplicação dos critérios deve ser realizada, preferencialmente, por duas ou mais pessoas, de forma independente, de modo a aumentar a confiabilidade dos resultados (DYBA *et al.*, 2007). Em caso de discordâncias quanto à inclusão ou exclusão de estudos, deve-se alcançar um consenso entre eles quanto à decisão e, caso persista o impasse, um terceiro revisor deve ser consultado. Cabe destacar a importância das decisões quanto a inclusão e exclusão de estudos. Normalmente, em revisões sistemáticas, apenas os artigos excluídos são alvo de revisões quanto às decisões. Esse equívoco é decorrente da suposição de que a inclusão de mais artigos não afeta negativamente o resultado da revisão (ao contrário da exclusão, em que os dados dos estudos selecionados são efetivamente perdidos). No entanto, a inclusão de mais estudos, potencialmente com resultados díspares (heterogêneos), prejudica a homogeneidade da revisão sistemática (MILLER, 2000), por seguinte enfraquecendo o valor do resultado final (principalmente quando aplicadas técnicas de meta-análise).

4.7. Avaliação da qualidade dos estudos primários

A determinação da qualidade do estudo primário é um importante instrumento para a posterior análise e síntese das evidências extraídas. Para esta etapa, devem ser definidos atributos e os critérios para sua

medição.

Os critérios de qualidade tem o propósito de caracterizar a qualidade do estudo em função de fatores rigor, credibilidade, apresentação, relevância dos resultados, dentro outros. Em tese, apenas a questão de rigor e credibilidade deveriam ser suficientes, dado que os estudos são, provavelmente, provenientes de eventos e publicações arbitrados. No entanto, a realidade é que vários trabalhos em Computação e, especificamente, Engenharia de Software, pecam nesse aspecto.

Duas abordagens podem ser utilizadas para a avaliação da qualidade: a análise individual dos artigos, com técnicas como a CASP, ou a análise da reputação e impacto do artigo. A primeira requer a leitura completa dos trabalhos. Embora seja uma técnica muito utilizada, o esforço necessário para sua aplicação é grande. A análise da reputação e o impacto do artigo, por outro lado, pode ser automatizada, viabilizando a avaliação de qualidade de muitos artigos. Observa-se ainda que ambas as abordagens podem ser combinadas (por exemplo, avaliando-se os artigos primeiramente pela reputação e, posteriormente, validar e ajustar a qualidade aferida com a CASP).

4.7.1. CASP

Um conjunto de critérios que pode ser utilizado é o *Critical Appraisal Skills Programme* (CASP)³ e princípios desejáveis em trabalhos de Engenharia de Software Experimental (KITCHENHAM *et al.*, 2002). Por exemplo, um possível conjunto de fatores e seus respectivos critérios é (DYBA *et al.*, 2007):

- **Rigor:** O método experimental é apropriado para a hipótese a que se destina o trabalho e ele foi aplicado corretamente?
 - Projeto experimental é apropriado para os objetivos da pesquisa.
 - Amostra utilizada e dos métodos utilizados para a obtenção dessa amostra são adequados.
 - Utilização de grupo de controle para controlar o tratamento.
 - Métodos de coleta de dados utilizados foram apropriados.
 - Métodos de análise de dados e os métodos para garantir que a análise foi realmente baseada dos dados são adequados.
- **Credibilidade:** Os resultados foram corretamente obtidos e apresentados?
 - Relação entre os pesquisadores e os participantes foi considerada.
 - Estudo provê os resultados com conclusões bem justificadas.
- **Relevância:** Os resultados são úteis (*i.e.*, contribuem para o estado da prática ou da arte)?

Para cada critério, deve-se determinar uma nota e, posteriormente, uma nota final que agregue a todos. Por exemplo, a nota de cada critério pode ser booleana (sim ou não) ou em uma escala mais ampla.

As notas atribuídas por cada revisor (cada estudo possui no mínimo duas notas) devem ser comparadas e um consenso deve ser alcançado quanto à nota final. O índice Kappa pode ser utilizado para verificar se os resultados atribuídos por cada revisor não foram significativamente diferentes: valores superiores a 0,80 são suficientes (LANDIS; KOCH, 1977).

4.7.2. Impacto

Uma abordagem complementar para a determinação da qualidade dos estudos primários é utilizar a reputação dos meios nos quais os estudos foram publicados. Assumindo-se que, na média, a qualidade das publicações é

³ <http://www.phru.nhs.uk/casp/casp.html>.

homogênea nos meios em que são disponibilizadas, é possível extrapolar que a qualidade do meio de publicação estabelece um patamar de qualidade para os estudos nele publicado.

Uma segunda extrapolação é que os autores que publicam esses estudos mantêm esse nível de qualidade em suas demais publicações. Finalmente, a terceira extrapolação: as instituições às quais esses autores estão vinculados também mantêm a qualidade de seus publicações em um nível homogêneo.

A cada extrapolação, a validade da afirmação tem menos força e risco de ser inválida. De fato, um mesmo autor pode ter trabalhos bons em publicações de impacto e estudos preliminares em publicações de impacto restrito e eventos regionais. Dessa forma, é necessário estabelecer um peso para cada nível de extrapolação, atribuindo maior valor para a publicação, um valor intermediário para o autor e um valor reduzido para a instituição. Além disso, é necessário esclarecer que tal critério, por ser de qualidade, não influencia na inclusão ou exclusão dos artigos a serem estudos, mas sim na confiança dos dados extraídos. De fato, ao invés de aferir a qualidade, pode-se utilizar esse critério apenas para classificar quais estudos serão lidos primeiro, de modo a otimizar a análise dos dados (embora todos os dados tenham de ser efetivamente lidos).

Algumas áreas tem sistematicamente coletado dados sobre o impacto de publicações, os autores que mais publicam nos meios de mais impacto e a instituições em que atual. Especificamente para Engenharia de Software, Uma série de estudos foram realizados para a caracterização dos principais autores e instituições da área (GLASS, 1994, 1995, 1996, 1998, 1999, 2000b, 2000a; GLASS; CHEN, 2001, 2002, 2003; TSE *et al.*, 2004; GLASS; CHEN, 2005; WONG *et al.*, 2008, 2009, 2011). Eles consideraram as principais revistas da área e contabilizaram os autores e instituições de seus artigos. As revistas analisadas foram:

- Information and Software Technology (IST), publicada pela Elsevier Science.
- Journal of Systems and Software (JSS), publicada pela Elsevier Science.
- IEEE Software (SW), publicada pela IEEE.
- Transactions on Software Engineering (TSE), publicada pela IEEE.
- Transactions on Software Engineering and Methodologies (TOSEM), publicada pela ACM.

Embora os autores influenciem, positivamente, a presença das instituições na classificação, eles não se configuram como um fator crítica. De fato, das 15 principais instituições de 2004, apenas três possuíam um dos autores que mais publicaram. De fato, é correto afirmar que, embora a influência dos principais autores seja relevante, a influência das instituições na quantidade de publicações feitas é maior (isso é uma característica diferente de outras áreas, tais como Sistemas de Informação (GLASS; CHEN, 2002)).

4.8. Extração de dados

Para a etapa de extração de dados, deve-se determinar quais dados serão extraídos de cada estudo selecionado e os procedimentos para sua extração.

Os dados a serem extraídos devem ser definidos e, preferencialmente, um formulário para a extração dos dados deve ser criado. Por exemplo, os seguintes dados geralmente devem ser extraídos dos estudos (DYBA *et al.*, 2007):

- tipo de estudo experimental (levantamento, estudo de caso, experimento),
- intervenção,
- amostra da população analisada,
- resultados obtidos.

Os procedimentos para a extração dos dados devem contemplar a extração de resultados objetivos e subjetivos:

- **Resultados objetivos:** Dados que podem ser diretamente extraídos do estudo. Por exemplo:
 - **Identificação:** Nome das publicações que referem-se a esse estudo, seus autores e a fonte da qual foi obtido.
 - **Metodologia:** Método utilizado para conduzir o estudo.
 - **Resultados:** Resultados do estudo.
 - **Problemas:** Limitações encontradas pelos autores do estudo.
- **Resultados subjetivos:** Dados que não podem ser extraídos diretamente do estudo. Por exemplo: informações sobre os autores, impressões gerais e abstrações sobre o estudo.

Tal como na seleção de estudos, a extração de dados deve ser realizada, preferencialmente, por duas ou mais pessoas, de forma independente, de modo a aumentar a confiabilidade dos resultados (DYBA *et al.*, 2007).

Além da extração dos dados, é necessário verificar a necessidade de preparar os dados para análise. Por exemplo, normaliza-los, colocando-os todos na mesma unidade de medida, permitindo que os resultados obtidos de cada estudo sejam comparáveis (BIOLCHINI *et al.*, 2005). No entanto, em Computação e, mais especificamente, em Engenharia de Software, os dados geralmente são incompatíveis, seja por questões do tipo do dado (escalar, faixa, etc), de erros na coleta dos dados (por exemplo, não ser estabelecida a precisão da medição e uso inconsistente de escalas em um mesmo estudo) ou até mesmo porque os dados são incomparáveis (por exemplo, a medida de linhas de código pode variar de acordo com a implementação da contagem, a linguagem da aplicação sendo medida, etc).

Como avaliar se os dados extraídos são suficientes para responder a questão de pesquisa estabelecida para a revisão sistemática? Da mesma forma com que critérios de seleção são estabelecidos e testados com *screening*, a extração de dados deve ser avaliada, extraindo-se os dados dos estudos selecionados do grupo de controle.

4.9. Análise dos resultados

Após a extração dos dados, eles devem ser analisados e devidamente apresentados. A disponibilidade de técnicas para a análise dos dados depende das características da revisão sistemática e dos estudos por ela analisados. As principais técnicas disponíveis são:

- **Meta-análise:** Método de análise estatística para combinar dados quantitativos.
- **Contagem de votos:**
- **Revisões tradicionais:**

Outras técnicas disponíveis, segundo (DIXON-WOODS *et al.*, 2005), são resumos narrativos, análise temática, *grounded theory*, meta-etnografia, meta-estudo, síntese realista, técnicas de análises de dados de Miles e Huberman, análise de conteúdo, levantamento de casos, análises comparativas e qualitativas, meta-análise Bayesiana. Cada uma dessas técnicas possui níveis diferentes de força (confiabilidade) dos resultados e pontos fracos, adequação ao tipo de evidências (quantitativas ou qualitativas) a serem analisadas e ao tipo de questão de pesquisa a ser respondida (DYBA *et al.*, 2007).

Cada técnica confere um nível distinto de confiabilidade aos resultados obtidos em sua aplicação (DYBA *et al.*, 2007) e, proporcional à essa confiabilidade, um conjunto de requisitos. A listagem acima está classificada em ordem decrescente de qualidade. A meta-análise é a técnica mais adequada para a obtenção de conclusões

confiáveis. No entanto, ele impõe severos requisitos quanto às evidências e aos respectivos estudos utilizados, especialmente quanto às variações entre os vários componentes do experimento (MILLER, 2000).

A heterogeneidade dos resultados dos estudos experimentais é uma limitação para a aplicação de procedimentos de meta-análise. O protocolo estabelecido durante o planejamento, os critérios de inclusão e de exclusão, a avaliação da qualidade dos estudos analisados, dentre outros possíveis mecanismos (MILLER, 2000) propiciam a redução dessa heterogeneidade, de forma que as técnicas de meta-análise possam se utilizadas. Obviamente, a sensibilidade dos resultados deve ser sempre verificada, utilizando-se, se possível, mais de uma técnica.

Observa-se, entretanto, que o emprego de meta-análise para estudos não-experimentais (algo não raro em revisões sistemáticas ou mapeamentos sistemáticos) é questionável e alvo de poucos estudos (MILLER, 2000): tais estudos não possuem mecanismos adequados para isenção de viés. Dessa forma, se estudos não experimentais são alvos da revisão sistemática, a técnica de meta-análise não é recomendada. A mesma recomendação é aplicável para revisões sistemáticas com estudos experimentais com resultados conflitantes: meta-análise tem o propósito de combinar resultados de experimentos, cada qual com poder estatístico insuficiente para aceitar ou rejeitar a hipótese nula; e não para analisar respostas díspares e estudos heterogêneos (MILLER, 2000).

A técnica de contagem de votos combina os resultados dos testes, classificando-os em categorias (*e.g.*, efeito positivo significativo, efeito negativo significativo, sem efeito significativo).

Os resultados da análise podem ser apresentados de diversas formas: tabelas, gráficos, etc.

As técnicas e procedimentos utilizados para analisar a sensibilidade dos dados devem ser apresentados. Comentários finais, resumindo todo os resultados da revisão sistemática, devem ser apresentados:

- Quantidade de estudos: Quantidade de estudos analisados.
- Análise de tendências/viés: O viés da busca e seleção de estudos e da extração de dados deve ser descrita.
- Discordâncias entre os revisores: Conflitos de opiniões dos revisores e os métodos para resolução dos conflitos devem ser descritos.
- Aplicação dos resultados: Sugestões de como os resultados da revisão sistemática podem ser aplicados.
- Recomendações: Recomendações dos revisores sobre a maneira como a revisão sistemática pode ser conduzida.

4.10. Considerações finais

Uma questão importante para o sucesso de uma revisão sistemática é a recuperação de todos as publicações e evidências relevantes e relacionadas à questão de pesquisa. O protocolo é o elemento da revisão sistemática que assegura que (1) serão obtidos todos os estudos relevantes, (2) que as evidências serão devidamente coletadas, (3) analisadas e sintetizadas.

Infelizmente, em Computação, atender a todos esses requisitos não é trivial, conforme retratado na Figura 4.2. Considere as evidências (**Evidence**). O objetivo de uma revisão sistemática é extrair as evidências dos estudos primários (**Study**), analisá-las e sintetizá-las, criando, dessa forma, novas evidências. Toda evidência é gerada por um estudo primário, sem exceção (alguns estudos, infelizmente, não geram evidência alguma, conforme já argumentado). Publicações relatam os estudos, descrevendo sua motivação, método e resultados, o que inclui as evidências. Aqui temos o primeiro problema: nem toda publicação está disponível em um meio apropriado: bibliotecas (**Library**) e bases bibliográficas (**Electronic database**). Dessa forma, virtualmente o estudo e suas evidências não existem.

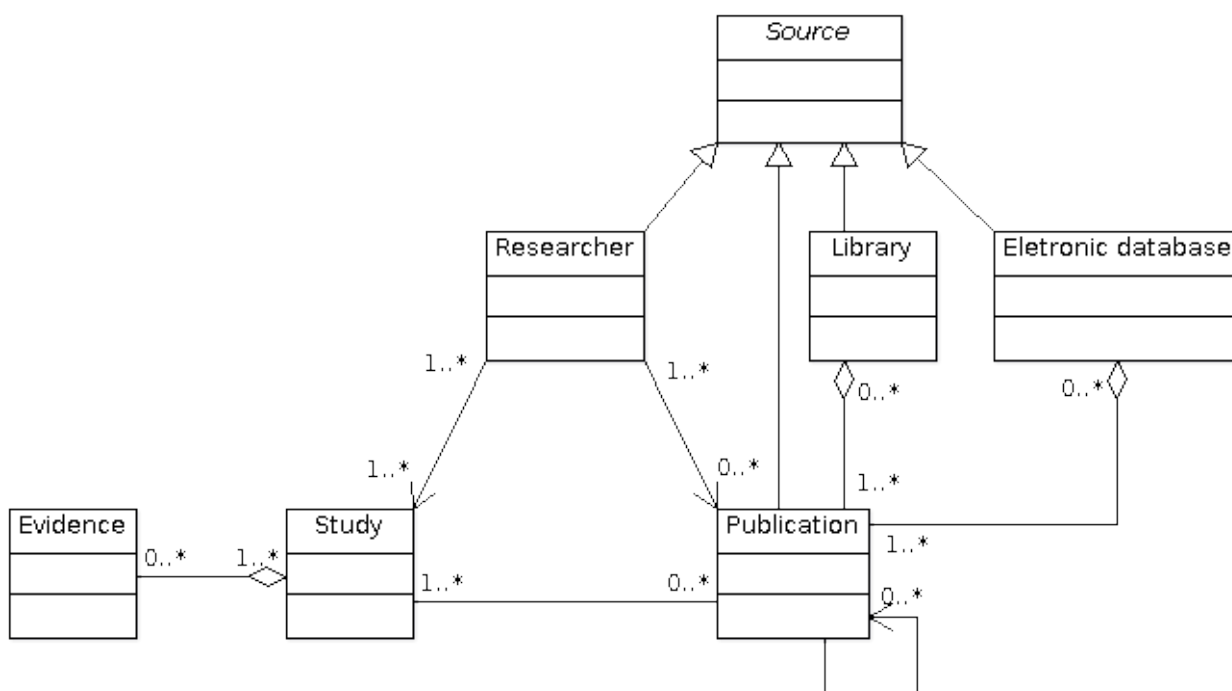


Figura 4.2: Evidências e seu relacionamento com estudos e fontes.

No entanto, ignorar as evidências não publicadas não é desejável: talvez elas sejam relevantes. Geralmente as publicações estão disponíveis, mas em locais pouco usuais (GLASS, 1994). Nesse caso, o único elo que existe entre o estudo e a publicação que permite a recuperação dela é o pesquisador (**Researcher**), participante do estudo e provável autor de publicações a respeito. Por esse motivo que especialistas da área sempre devem ser consultados durante a revisão sistemática: eles são, em última instância, o único elo que existe entre um estudo e a publicação de suas evidências.

Considerando-se que ao menos uma publicação tenha sido encontrada sobre a questão em pesquisa, pode-se utilizar a própria publicação como fonte. Embora se evite o uso das publicações como fonte, isso é necessário quando (1) são encontradas poucas publicações e os especialistas não conseguem sugerir outros estudos relevantes ou (2) devido a limitações nos mecanismos de pesquisa das bibliotecas e bases eletrônicas.

Quando utilizadas fontes de estudo baseadas em mecanismos de busca, observa-se uma limitação quanto à precisão da recuperação por expressão de busca. Apesar da técnica de revisão sistemática ditar que a expressão é diretamente obtida dos atributos da questão (PICO), o fato é que, em Computação, essa medida é insuficiente para a obtenção de resultados satisfatórios.

As buscas, realizadas geralmente em resumos, não recuperam todos os trabalhos relevantes. Isso decorre do fato de que os resumos de computação carecem de informações relevantes para a apropriada caracterização dos trabalhos e de seus resultados (BUDGEN *et al.*, 2008; KITCHENHAM *et al.*, 2008). Ao mesmo tempo, trabalhos de baixa qualidade, mas que incluem os termos da expressão de busca, são encontrados. Por exemplo, várias bases de busca não utilizam uma base controlada de palavras-chaves ou permitem a definição de palavras-chaves sem suficiente critério (sem limitação quanto a quantidade de palavras-chave, sem verificar se as palavras escolhidas realmente condizem com o resumo, etc).

Uma medida que pode ser tomada é a verificação, *a posteriori*, dos resultados obtidos pelas buscas, o que reduz o problema dos trabalhos irrelevantes encontrados (embora esta solução demande um maior esforço por parte dos pesquisadores que conduzem a revisão sistemática). No entanto, quanto aos trabalhos relevantes (que não foram encontrados), restam duas alternativas:

- Adicionar termos que permitam a seleção dos estudos relevantes.
- Adicionar os estudos por outros meios (ou seja, acrescentar novas fontes de estudos).

A inclusão dos estudos como controle é a solução normalmente utilizada. No entanto, a inclusão indiscriminada desses, sem a aplicação dos mesmos critérios aplicados a todos os outros estudos, cria um viés indesejado na revisão. A solução adequada seria a alteração dos termos da expressão de busca e, se necessário, corrigir os atributos da questão e até mesmo a questão.

A alteração dos termos em função dos itens de controle acrescenta, no entanto, duas novas categorias de problemas: o aumento da quantidade de estudos a serem analisados (se os termos forem ampliados) ou a redução excessiva do escopo da revisão (em função da seletividade dos estudos de controle).

Uma alternativa para controlar esse problema é o estabelecimento de requisitos quanto a precisão (*precision*) e a recuperação (*recall*) da busca em função do grupo de controle de estudos. Dessa forma, espera-se que a busca possua uma recuperação elevada (recupere uma porcentagem significativa dos estudos do grupo de controle) e uma precisão razoável (não excessivamente alta, de modo que apenas itens do grupo de controle sejam encontrados, e nem baixas demais, de modo a encontrar muitos itens irrelevantes).

Outra alternativa é a utilização dos próprios estudos encontrados como fontes de estudos, observando-se os demais trabalhos por eles citados.

Seleção de fontes

A seleção das fontes de estudo tem início pela criação de um grupo de controle de fontes definidos pelos executores da revisão sistemática e pelos pesquisadores externos. Essas fontes são avaliados quanto aos critérios de seleção de fontes e aquelas aprovadas são utilizadas para como fontes. Outras fontes, comuns à maioria das revisões sistemáticas, são também analisadas. Para todas as fontes selecionadas, são então definidos os procedimentos de busca (quase sempre baseado em expressões de busca).

5.1. Avaliação do grupo de controle

O grupo de controle é formado pelas fontes de estudos sugeridas pelos pesquisadores (geralmente pelos pesquisadores externos). Os critérios de seleção de fontes são então aplicados a essas fontes e se determina o conjunto de fontes selecionadas.

As principais fontes de sugestões sugeridas são revistas e eventos científicos. Eventualmente, pode-se também obter as fontes a partir dos estudos definidos no estudo de controle, identificando-se os locais em que eles foram publicados.

5.2. Seleção das fontes

As fontes do grupo de controle já fazem parte do grupo de fontes selecionadas. Porém, nem sempre é possível analisar diretamente tais fontes, sendo necessário encontrar um mecanismo de indexação e busca de publicações que contemple as fontes do grupo de controle.

As publicações realizadas em revistas (*journals*) e eventos (simpósios, congressos, grupos de trabalho, dentre outros publicados em anais – *proceedings*) são, em sua maioria, disponibilizadas em uma base de publicações. Por exemplo, eventos sob o patrocínio da ACM são armazenados e indexados pela ACM Digital Library. Portanto, ao invés de pesquisar diretamente na fonte (a revista ou nos anais de um evento), é possível realizar a busca na base de dados (*bibliographic database* ou em algum serviço que indexe essas

bases de dados (*indexing service*).

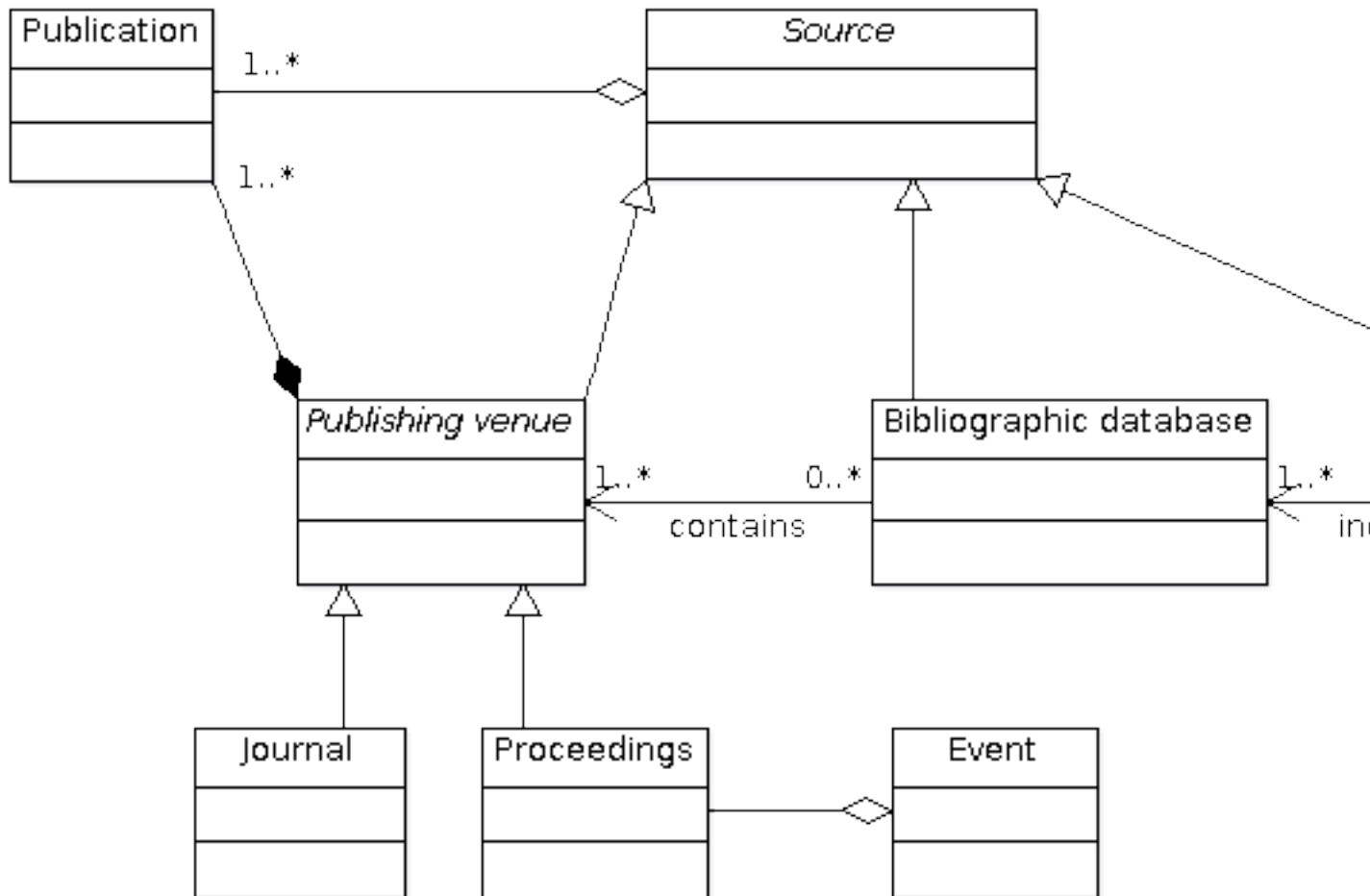


Figura 5.1: Fontes de estudos.

Além disso, outras fontes de estudo, tradicionais para a área em se tratando de revisões de literatura, também devem ser consideradas para revisões sistemáticas. Em Engenharia de Software, as seguintes fontes devem ser consideradas:

- IEEE Xplore Digital Library (<http://ieeexplore.ieee.org/>),
- ACM Digital Library (<http://portal.acm.org/>),
- Scirus (<http://www.scirus.com/>),
- SpringerLink (<http://www.springerlink.com/>),

- Google (<http://www.google.com>),
- Google Scholar (<http://scholar.google.com>),
- ISI Web of Knowledge (<http://portal.isiknowledge.com/>),
- ISO (<http://www.iso.org>),
- INSPEC (<http://www.iee.org/publish/inspec/>),
- Management Science (JSTOR) (<http://www.jstor.org/search/>),
- CiteSeer (<http://citeseer.ist.psu.edu/>),
- The Collection of Computer Science Bibliographies (<http://liinwww.ira.uka.de/bibliography/>),
- Digital Bibliography and Library Project (DBPL) (<http://dblp.uni-trier.de/>),
- CiteSeer (<http://citeseer.ist.psu.edu/>),
- io-port.net (http://io-port.net/index_eng.html),
- SciVerse Scopus (<http://www.scopus.com>).

A utilização de todas essas fontes é inviável atualmente, dado que todas definem interfaces distintas de acesso (e não programáveis). Logo, deve-se escolher as mais relevantes para o tema em pesquisa. Além disso, a preferência da seleção de fontes deve recair nos mecanismos mais abrangentes e confiáveis. Observando-se a Figura 5.1, os serviços de indexação, que abrangem mais de uma base bibliográfica (e, por conseguinte, mais de um veículo de publicação) possuem a preferência de escolha. Por exemplo, a ACM Digital Library, que contém a base bibliográfica da ACM, também indexa a base bibliográfica da IEEE. Assim, a utilização do mecanismo de busca da ACM elimina a necessidade de busca direta na base bibliográfica da IEEE (já que o serviço de indexação do ACM DL já o faz). O Scopus indexa a maioria das revistas publicadas, mas não indexa publicações em conferências. Se o critério de seleção de fontes restringir os resultados àqueles publicados em revistas, o Scopus é suficiente para uma revisão sistemática.

5.3. Avaliação das fontes

O método de busca baseado em expressões e as bibliotecas digitais indexadas possuem algumas limitações. Ferrari e Maldonado (2008), after facing some issues with such applications, created a classification based upon features that are rather useful for systematic reviews:

- **Centralized results:** Centralization is defined regarding the databases used by the search engine. When the data belongs to the same organization that provides the search engine, it is called centralized. If the database belongs to another company and just the metadata is locally available, it is not considered centralized.
This is important as systematic reviews requires the full text of the studies, which are only available on search engines with centralized results. Another important fact is that, as the database and the search engine is tightly integrated, the probability that invalid data (such as incomplete abstracts, incorrect page information or conference naming) will be provided as result is reduced.
- **Full string composition:** This feature is related to the capability the search engine has of processing complex and long search expressions. The search expression complexity is measured by the use logical operators (such as *or*, *and* and *not*). As some search engines have restrictions upon the length of the expressions, this is also accounted.
- **Adequate scope narrowing:** A search engine with adequate scope narrowing allows the specification the fields (abstract, keyword, or full text) that will be scanned.
Some search engines do not allow scope narrowing (*e.g.*, Google Scholar). Others provides scope narrowing, but it does not work property (*e.g.*, ACM, which search on keywords even when only

abstracts are selected).

- **Consistent results:** A result is given as consistent when every run with the search expression results in the same set of publications and when composed search terms are treated as atomic terms (e.g., “*term1 term2*” is different than “*term1*” and “*term2*”).

5.3.1. Caracterização das fontes

In order to correctly assess the search engines and the threads they pose to the systematic review, every source that has been selected must be tested and classified against these characteristics. The following procedures are recommended:

- Além da busca realizada pela expressão na forma conjuntiva (CNF), devem ser realizadas buscas com as cláusulas resultantes da transformação da expressão da forma conjuntiva para disjuntiva (DNF). O resultado final de ambas as buscas deve ser exatamente o mesmo. Se não o for, o resultado a ser considerado é o das buscas realizadas com as cláusulas (e não a busca realizada com a expressão CNF).
- Uma amostra aleatória deve ser obtida dos resultados obtidos. Desta amostra, deve-se verificar se todos os elementos atendem aos requisitos de todas as expressões de busca utilizada para obtê-lo. Se um estudo, encontrado por duas ou mais expressões de busca, não atende os argumentos de todas as suas expressões (mas atende a pelo menos uma das expressões), os resultados podem ser considerados, mas existe uma ameaça real à validade do estudo. Se um estudo encontrado por alguma expressão de busca não atende a nenhuma das expressões de busca que o detectaram, todos os estudos relativos ao mecanismo de busca devem ser desconsiderados.

No caso de uma revisão sistemática iterativa, pode-se utilizar a primeira iteração para o ensaio e ajuste da expressão de busca. A partir de um conjunto de estudos, estabelecidos a priori por especialistas da área, o conjunto de estudos encontrados pelo mecanismos de busca será aferido quanto a sua precisão (*precision*) e cobertura (*recall*). As expressões de busca devem ser ajustadas, com devidas justificativas e análises quanto a inserção de viés, para alcançar o máximo de cobertura e o mínimo de cobertura (sem a exclusão de termos obrigatórios da revisão sistemática).

5.4. Definição de expressões de busca específicas

Devido à falta de padronização entre os mecanismos de busca disponíveis, é necessária a definição de uma expressão de busca específica para cada fonte de estudo selecionado (DYBA *et al.*, 2007). Nesta seção, instruções para a construção das expressões específicas, a partir da expressão original, são estabelecidas.

Cabe ressaltar que as expressões resultantes não são necessariamente equivalentes à expressão original (devido a limitações das fontes de estudo). Embora no momento da construção deste relatório, tais problemas tenham sido identificados e devidamente documentados, as constantes alterações dos mecanismos de busca (FERRARI; MALDONADO, 2008) podem acarretar em comportamentos diferentes do aqui previsto. Portanto, é desejável que mecanismos sejam estabelecidos para assegurar que os resultados obtidos estão corretos (como, por exemplo, o prévio estabelecimento de um grupo de artigos que sevirá de controle, *i.e.*, que deve ser encontrado pelos mecanismos de busca).

Os mecanismos de busca os quais foram utilizados para esses procedimentos são todos sites Web. O seu acesso requer um navegador comum. No entanto, para obter os estudos e armazená-los localmente, é possível utilizar extensões ou *plugins*, de modo a agilizar o processo. Uma das ferramentas recomendadas, para os usuários do navegador Firefox (Mozilla Foundation, 2004), é o DownloadThemAll! ([http:](http://)

`//www.downdthemall.net/`). Ele permite a obtenção de todos os arquivos disponíveis (ligados) em uma página Web. Ele ainda possibilita escolher o tipo dos arquivos que serão obtidos automaticamente, seja pela extensão dos arquivos ou de expressões regulares, a geração de nomes de arquivo (alguns mecanismos de busca – *e.g.*, SpringerLink – usam sempre o mesmo nome de arquivo – *e.g.*, `fulltext.pdf` –, o que é um empecilho para o gerenciamento dos estudos recuperados. Ainda para o Firefox, recomenda-se a instalação da extensão Zotero ((Center for History and New Media (George Mason University), 2006)), a qual permite a extração de dados bibliográficos de uma página Web (algo extremamente útil para mecanismos que não exportam os dados bibliográficos de várias publicações ao mesmo tempo, tal como a ACM Digital Library).

Para algumas fontes, são documentadas várias formas de busca, sempre em função do período em que tais procedimentos eram válidos. Por exemplo, a ACM DL mudou seu mecanismos de busca em 2010, inutilizando os procedimentos anteriores. Para fins de documentação e história, tais procedimentos, ainda que obsoletos, são mantidos neste relatório.

5.4.0.1. ACM Digital Library (antes de 2010)

A página de busca do ACM Digital Library é <http://portal.acm.org/advsearch.cfm>. O mecanismo de busca disponibilizado não permite a utilização direta de operadores booleanos.

Além da expressão de busca, deve-se selecionar a biblioteca de publicações a ser utilizada (opção disponível no topo da página). Certifique-se de selecionar *The ACM Digital Library* (e não a opção padrão, *The Guide*).

Outra opção a ser definida é quanto ao conteúdo disponível das publicações. Essa opção encontra-se no final do formulário, em *Results must be accessible*. Deve-se selecionar apenas o item *Full Text* (textos completos).

A expressão de busca é uma composição de termos. Para cada termo, pode-se especificar sua obrigatoriedade quanto a presença nos resultados: obrigatório, opcional, proibido. Para obrigar a presença de um termo, prefixe-o com o símbolo `+`. Para indicar como opcional, não é necessário acrescentar símbolo algum. Para proibir um termo, utilize o símbolo `-`. Exemplos:

- Para buscar artigos que tenham sempre os termos `testing` e `component`: `+testing +component`.
- Para buscar artigos que tenham sempre o termo `testing` e, opcionalmente, `component`: `+testing component`.
- Para buscar artigos que tenham sempre o termo `testing` e não contenham o termo `component`: `+testing -component`.

Termos de busca com várias palavras devem ser protegidos com aspas (por exemplo, `"quality standard"`).

Para buscar variações de um termo, pode-se empregar coringas (*wildcards*). A ACM Digital Library suporta apenas um coringa, o asterisco (`*`). Ele pode ser substituído por um ou mais caracteres quaisquer.

A busca, por padrão, é realizada no texto completo das publicações. Para restringí-la a uma parte específica, especifique a zona a ser pesquisada, prefixando o termo de busca com o nome da zona e o símbolo `:`. Por exemplo:

- Para buscar artigos que tenham sempre os termos `testing` e `component` no resumo: `+abstract:testing +abstract:component`.
- Para buscar artigos que tenham sempre o termo `testing` e, opcionalmente, `component` no resumo: `+abstract:testing +abstract:component`.

- Para buscar artigos que tenham sempre o termo `testing` no resumo e não contenham o termo `component` no corpo do documento: `+abstract:testing -component`.

Em tese, as expressões de busca com vários termos, se não especificados modificadores como o `+` e `-`, funciona como uma expressão `OR`, ou seja, todas as publicações que atendam a algum dos termos serão recuperadas. Infelizmente não é isso que ocorre na prática: todos os termos são unidos com o operador `AND`. Assim, para obter o efeito `OR`, é necessário realizar várias buscas, cada qual com uma expressão `AND`. O efeito `OR` é obtido posteriormente com a junção, realizada manualmente, dos resultados das consultas `AND`.

Para as páginas de resultados de busca da ACM Digital Library, todos as referências aos arquivos dos artigos estão no formato `http://portal.acm.org/ft_gateway.cfm?id=* & type=pdf`. Esse padrão pode ser utilizado para a obtenção automatizada dos arquivos completos das publicações.

5.4.0.2. CiteSeer Library

5.4.0.3. IBM Journal

5.4.0.4. IEEE Xplore

O endereço do mecanismo de busca da IEEE, o IEEE Xplore, encontra-se em `http://ieeexplore.ieee.org/search/advsearch.jsp`. Ele indexa apenas artigos em inglês.

As buscas no IEEE Xplore, quando inseridas no formato livre (*Free-Form Advanced Search*) são uma seqüência de expressões de buscas combinadas com operadores. Cada expressão de busca é colocada entre parênteses. Por exemplo, para pesquisar artigos que contenham as palavras `quality standard` ou `quality model` e, também, as palavras `software` ou `computer program`, a seguinte expressão de busca poderia ser utilizada:

```
(quality standard <or> quality model) <and> (software <or> computer program)
```

Os operadores dessa expressão estão no formato `infix`, o que torna a escrita das expressões mais simples. No entanto, é possível escrever as expressões no formato `prefix`. Para expressões mais complexas, esse formato é interessante, dado que vários operadores (como o `<or>` e o `<and>` suportam vários operandos (e não apenas dois, ao qual o IEEE Xplore restringe quando utilizando o formato `infix`). Por exemplo, a expressão anterior pode ser reescrita como:

```
(<or>(quality standard , quality model)) <and> (<or>(software , computer program))
```

Ou, caso prefira uma expressão totalmente prefixa:

```
<and>(((<or>(quality standard , quality model)), (<or>(software , computer program))))
```

Os operadores de composição do IEEE Xplore funcionam como chamada de funções. Cada identificador de função é imediatamente precedido pelo símbolo `<` e sucedido pelo símbolo `>`. Os argumentos de cada função são colocados entre parênteses e são separados por vírgula. Não é necessário colocar o texto de um argumento, composto por várias palavras, entre aspas. Por exemplo:

```
<or>(quality standard , maturity model , quality factor , quality attribute)
```

Os operadores suportados pelo mecanismo de busca são¹:

- **<and>**: Verifica a presença de todos os operandos.
- **<or>**: Verifica a presença de ao menos um dos operandos
- **<not>**: Verifica a ausência de todos os operandos.
- **<in>**: Verifica a existência de todos os operandos em um determinado campo. Exemplos de campos e respectivas palavras-chaves: resumo (**ab**), título (**ti**).
- **<paragraph>**: Verifica a existência de todos os operandos no mesmo parágrafo.
- **<sentence>**: Verifica a existência de todos os operandos na mesma oração.
- **<phrase>**: Verifica a existência de todos os operandos na mesma frase.

Os termos de busca podem conter coringas (*wildcards*), permitindo assim a busca por variações das palavras. Os coringas disponíveis são:

- **?**: Representa um único caractere alfa-numérico na posição atual. Pode ser colocado em qualquer local da palavra (início, meio e fim).
- *****: Representa um ou mais caracteres alfa-numéricos na posição atual. Pode ser colocada apenas no meio ou final de uma palavra (não pode ser colocado no início de uma palavra).

Importante lembrar que o uso de coringas desativa a busca por variações comuns de uma palavra (*stemming*).

Outro recurso disponibilizado pelo IEEE Xplore é a busca por palavras exatas. Nesse caso, elas devem ser apresentadas entre aspas (por exemplo, "cable"). Além disso, as seguintes restrições se aplicam:

- Sinais de pontuação na palavra devem ser substituídos por espaços em branco.
- Caracteres coringas não podem ser utilizados.

Em tese, as expressões de busca do IEEE Xplore são muito versáteis. No entanto, o mecanismo de execução não parece suportar todas as opções documentadas corretamente. Por exemplo, as expressões abaixo, apesar de equivalentes, apresentam resultados totalmente diferentes:

```
(quality standard <or> quality model) <and> (software <or> computer program)
<or>(quality standard , quality model)) <and> (<or>(software , computer program))
<and>((<or>(quality standard , quality model)), (<or>(software , computer program)))
```

A primeira apresenta os resultados corretos. A segunda não apresenta resultado algum: o mecanismo de busca, por algum motivo, ignora as vírgulas como separador de argumentos e procura as palavras `quality standard quality model` e `software computer program` (e a chance de algum documento conter tais palavras inteiras é realmente baixa). A terceira expressão não é executada: o IEEE Xplore não a considera uma expressão válida.

Assim, recomenda-se a utilização do primeiro formato de expressões de busca, com um operador para cada dois operandos.

Estabelecidas as expressões de busca, pode-se proceder à busca propriamente dita. A página de busca do IEEE Xplore (<http://ieeexplore.ieee.org/search/advsearch.jsp>) oferece várias opções. Os ajustes realizados para elas concordarem com os critérios estabelecidos foram:

¹ a lista completa e de operadores encontra-se em http://ieeexplore.ieee.org/xplorehelp/Help_start.html#Help_searchoperators.html

- Publicações (*Publications*): Todas as publicações disponíveis devem ser selecionadas (*IEEE Periodicals, IET Periodicals, IEEE Conference Proceedings, IET Conference Proceedings, IEEE Standards*).
- Outros recursos (*Other Resources*): Não deve ser selecionado.
- Estado do padrão (*Standard Status*): Selecionar todos os padrões, independentemente do estado (*All*).
- Data (*Select date range*): Não impor limitações quanto à data. Para isto, deve-se selecionar as publicações de todos os anos (*From year All*) até a data presente (*to Present*).
- Formato de apresentação (*Display Format*): Escolher a apresentação da referência do artigo apenas (*Citation*). Os resumos serão obtidos posteriormente (e os resumos apresentados com essa opção possuem limite de tamanho, diminuindo a sua utilidade).
- Organização dos resultados (*Organize results*): O limite de resultados encontrados (parâmetro *Maximum*) deve ser alterado para 500 (o maior valor permitido). O limite de resultados a serem apresentados a cada tela (*Display results per page*) deve ser configurado para 100 (o maior valor permitido). Os resultados devem ser ordenados (*Sort by*) por ano (*Year*) e em ordem (*in order*) decrescente (*descending*).

O mecanismo de busca da IEEE ainda pode ser manipulado para facilitar a recuperação dos artigos identificados na busca. Apesar do limite de artigos, no formulário, ser 500, e do máximo visualizável por páginas ser 100, é possível alterar ambos os parâmetros diretamente no endereço Web (URL) gerada ao submeter o formulário.

O conceito é bem simples. Realizada a pesquisa (submissão do formulário), conforme descrito nos passos anteriores, o endereço da página Web dos resultados contém todos os parâmetros de busca que constavam no formulário. Para alterar os limites de artigos visualizáveis, é suficiente alterar os parâmetros `ResultCount` e `maxdoc`. Eles representam, respectivamente, a quantidade de artigos mostrados por página e quantidade máxima de artigos recuperáveis. Por exemplo, pode-se utilizar como valores de ambos a quantidade de publicações encontradas (valor encontrado na página de resultado da busca, na frase *Your search matched XXX of XXXXX documents.*). Alteração realizada, basta carregar o novo endereço Web para obter o conteúdo em uma única página.

Para as páginas de resultados de busca do IEEEExplore, todos as referências aos arquivos dos artigos estão no formato `http://ieeexplore.ieee.org/iel*.pdf`. Esse padrão pode ser utilizado para a obtenção automatizada dos arquivos completos das publicações.

Também é possível obter o texto da entrada BibTex da publicação. Para cada publicação desejada, deve-se acessar a referência *AbstractPlus*. Na nova página, no lado esquerdo, existe um formulário para obter os metadados da publicação (*Download this citation*). Selecionado o formato (*Choose*) desejado e opção *BibTex* para *download*, basta submeter o formulário (acionamento do botão *Download*) para obter os metadados no formato BibTex.

Uma maneira alternativa para obter o metadado de todos os artigos é, na página de resultados da busca, selecionar todos os artigos (acionando-se a referência *Select All*) e acionar o botão *view selected items*. Na nova página, basta proceder como descrito no parágrafo anterior para obter os metadados de todos os artigos.

5.4.0.5. Inspec

5.4.0.6. Scirus (Elsevier)

A página de busca do Scirus simples é `http://www.scirus.com/srsapp/` e a página de busca avançada é `http://www.scirus.com/srsapp/advanced/index.jsp`.

O Scirus permite restringir o tipo de publicação a ser pesquisada (campo *Information types*). Usualmente este campo deve ser restrito a publicações em revistas (*Articles*), conferências (*Conferences*) e livros (*Book*).

No campo *Content Sources*, desabilite a pesquisa na Web como um todo (item no rodapé do campo, *The rest of the scientific web*).

A página de preferências do Scirus (*Search Preferences*) permite a especificação da quantidade de itens encontrados a serem mostrados a cada página (*Number of results*). Selecione o valor máximo (100) e salve as preferências (*Save Preferences*).

Aparentemente, ao realizar a busca a partir da página avançada, o Scirus desabilita a busca nas bases Web preferidas (*Preferred Web source*). Para ter certeza que a busca foi realizada corretamente, solicite uma nova busca a partir da página de resultados.

A expressão de busca é uma composição de termos com operadores booleanos e (AND), ou (OR) e não (ANDNOT). Por exemplo:

```
("quality standard" OR "maturity model") AND "software architecture"
```

Para as páginas de resultados de busca do Scirus, todos as referências aos arquivos dos artigos estão no formato http://www.scirus.com/srsapp/sciruslink?src=*. Esse padrão pode ser utilizado para a obtenção automatizada dos arquivos completos das publicações.

5.4.0.7. Compendex

5.4.0.8. ScienceDirect

5.4.0.9. Biblioteca da Keele University

5.4.0.10. Biblioteca Digital Brasileira de Computação

5.4.0.11. SpringerLink

A página de busca do SpringerLink é <http://www.springerlink.com/home/main.mpx>. Ela apresenta algumas limitações (FERRARI; MALDONADO, 2008):

- A opção de “Pesquisa Avançada” permite criar expressões de busca utilizando operadores lógicos para combinar os termos desejados. Entretanto, existem limitações para a formação das expressões, sendo que expressões de busca extensas (por exemplo, como a utilizada na busca da máquina do IEEE) não retornam os resultados desejados.
- A procura por termos compostos (por exemplo, ‘ ‘**aspect-oriented**’ ’) não funciona como desejado. A busca retorna publicações que possuem todas as partes do termo composto, mas não necessariamente juntos.
- A princípio, a máquina permite restringir a busca aos títulos e resumos das publicações, da mesma forma que a máquina do IEEE. Entretanto, observou-se, com algumas tentativas de busca, que a restrição não é atendida e a busca é feita no corpo do texto, retornando uma grande quantidade de publicações sendo que diversas delas não eram relevantes para esta revisão.

Para amenizar as dificuldades e manter a Springer como um dos repositórios pesquisados, deve-se adotar uma estratégia semelhante à utilizada para a busca na máquina da ACM:

- Construir expressões de busca considerando a combinação dos termos mais relevantes para as questões de pesquisa.
- Selecionar todos os periódicos/publicações para busca.
- Realizar a busca com as opções de pesquisa da aba “*Article by text*” da página de busca avançada. Os campos de pesquisa devem ser preenchidos da seguinte forma:
 - No campo “*Search for*”, inserir a expressão de busca.
 - Em “*Using*”, selecionar a opção “*All Words*”.
 - Em “*Within*”, marcar opção “*Abstract (Includes Title)*”.
 - Os demais campos devem permanecer na configuração padrão.

Para as páginas de resultados de busca do SpringerLink, todos as referências aos arquivos dos artigos estão no formato http://www.springerlink.com/content/*.pdf. Esse padrão pode ser utilizado para a obtenção automatizada dos arquivos completos das publicações.

5.4.0.12. Google Scholar

O endereço do mecanismo de busca do Google Scholar é <http://scholar.google.com>. O idioma de busca pode ser limitado a inglês e português (`lr=lang_en\%7Clang_pt`). A quantidade de resultados por página deve ser configurada para 100 (`num=100`). Deve ser habilitada a opção para mostrar os dados da publicação no formato BibTeX. Essas opções foram todas configuradas nas preferências do Google Scholar (http://scholar.google.com/scholar_preferences), as quais estão disponíveis para usuários cadastrados e autenticados no Google.

A busca deve ser restrita à área de computação (*Engineering, Computer Science, and Mathematics*). Essa opção é configurável na página de busca avançada do Google Scholar (http://scholar.google.com/advanced_scholar_search).

Um exemplo de expressão de busca no Google Scholar:

```
allintext: standard OR "maturity model" OR "meta-model" OR "quality assurance  
model" OR "quality factor" OR "quality attribute" quality product OR process  
OR "life cicle" software
```

5.4.0.13. ISO

O endereço da página de busca da ISO é http://www.iso.org/iso/en/Standards_Search.StandardsQueryForm. Os termos de busca utilizados foram:

```
("quality standard" OR "maturity model" OR "quality factor" OR "quality attribute")  
AND  
("software architecture")
```

5.4.0.14. ISI Web of Knowledge

A página de busca do ISI Web of Knowledge é http://portal.isiknowledge.com/portal.cgi?DestApp=XS_FORM&Func=Frame. Os termos de busca utilizados foram:

```
(standard OR "maturity model" OR "quality factor" OR "quality attribute")
AND
("software architecture")
```

5.4.0.15. JSTOR

O endereço da página de busca do *Journal Storage: The Scholarly Journal Archive* (JSTOR) é <http://www.jstor.org/search/> (busca básica). Para esta revisão, utilizou-se a opção de busca mais avançada disponível, localizada em <http://www.jstor.org/search/ExpertSearch>. A busca foi restringida para os idiomas português e inglês.

```
(standard OR "maturity model" OR "quality factor" OR "quality attribute")
AND
("software architecture")
```

5.4.1. CiteSeer.IST

O endereço da página de busca do CiteSeer.IST é <http://citeseer.ist.psu.edu/>. Os termos de busca utilizados foram:

```
software OR standard OR "product quality" OR "process quality"
```

5.4.1.1. The Collection of Computer Science Bibliographies

O endereço da página de busca do *The Collection of Computer Science Bibliographies* é <http://liinwww.ira.uka.de/bibliography/#search>.

As opções de busca foram configuradas para mostrar os resultados no formato BibTex, com 200 resultados por página, ordenados pelo ano e incluir na busca apenas os artigos disponíveis online.

```
(standard OR "maturity model" OR "quality factor" OR "quality attribute")
AND
("software architecture")
```

5.4.2. DBPL

O endereço de busca do DBPL é <http://www.informatik.uni-trier.de/~ley/db/indices/t-form.html>. Nesse caso, a busca realizada é por título. O DBPL não suporta operadores de combinação booleanos, assim foram realizadas várias buscas, cada qual com um conjunto das palavras-chave desejadas:

```
software quality standard
software quality model
software process quality standard
software process quality model
software product quality standard
software product quality model
```

5.4.3. io-port.net

Foi habilitada a opção *Use stemming (English) instead of truncation*. Os resultados foram ordenados por ano de publicação.

Busca avançada desabilitada para usuários sem autenticação.

5.4.3.1. Wiley InterScience

A página de busca da Wiley InterScience é <http://www3.interscience.wiley.com/search/allsearch>.

Seleção de estudos

A fase de seleção envolve a identificação e seleção das fontes e dos estudos primários.

1. Seleção e avaliação das fontes.
2. Seleção e avaliação dos estudos.

A partir da lista inicial de fontes de estudo, avalia-se se alguma nova fonte deve ser acrescentada. Essa nova lista é então avaliada de acordo com os critérios de seleção de fontes estabelecidos no protocolo. A remoção de fontes de estudo deve ser devidamente documentada, explicando o motivo da decisão. As fontes aprovadas e reprovadas devem então ser submetidas ao crivo dos especialistas, que detectará possíveis falhas no processo.

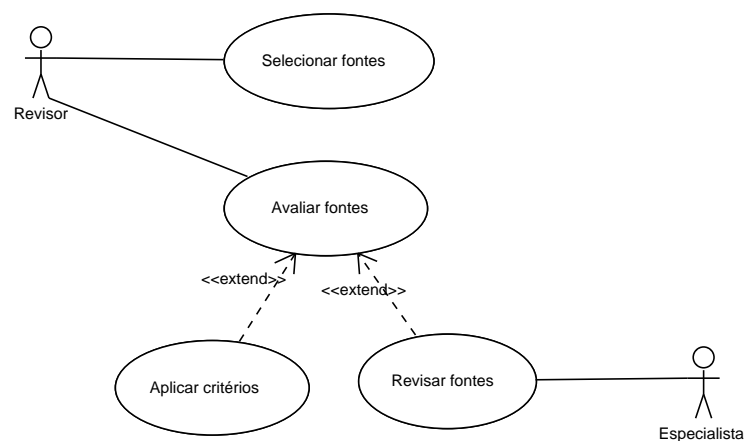


Figura 6.1: Caso de uso de seleção de fontes de estudos.

Escolhidas as fontes de estudo, procede-se a seleção de estudos. O método de seleção deve ser o determinado no protocolo e executado para cada fonte selecionada. Da mesma forma que as fontes, os

estudos identificados devem ser avaliados quanto aos critérios, no caso de inclusão e de exclusão, definidos no protocolo. O resultado da avaliação de cada estudo deve ser registrado e revisado pelo especialista, detectando possíveis falhas no processo.

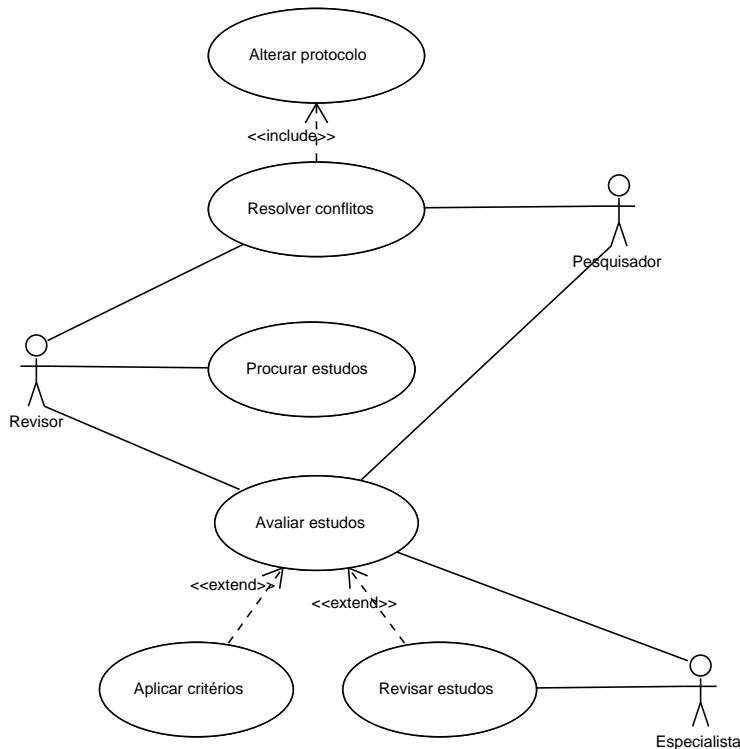


Figura 6.2: Caso de uso de seleção de estudos.

A seleção de estudos requer a obtenção das publicações referentes aos estudos identificados nas fontes de estudos. Geralmente são encontrados muitos trabalhos a serem analisados, o que inviabiliza a recuperação manual dos dados. Assim, para auxiliar a recuperação dos documentos, ferramentas devem ser utilizadas.

Uma classe de ferramenta recomendada para este fim são os gerenciadores de *downloads*. Especificamente para o navegador Firefox, recomenda-se o DownloadThemAll! (<http://www.downthemall.net/>). Ele permite a recuperação automática dos arquivos disponíveis em uma página Web, filtrando-os por uma expressão regular. Para cada fonte de estudos, uma expressão regular distinta é necessária.

Todos os arquivos obtidos, bem como a documentação da revisão sistemática, devem ser colocados sob controle de versão (*e.g.*, Subversion). Assim é possível controlar com maior precisão a execução da revisão sistemática.

Uma variação possível na seleção dos estudos, que garante uma melhor qualidade da seleção, é a avaliação de um subconjunto dos estudos, definido aleatoriamente, por um outro revisor ou pesquisador. Caso o resultado da avaliação seja distinto, deve-se resolver o conflito das avaliações e, caso confirmado o erro do protocolo, o mesmo deve ser revisado de modo a identificar o ponto que deve ser esclarecido para que conflitos semelhantes não ocorram nas próximas iterações. Caso seja confirmado um erro do revisor, deve-se revisar os demais estudos e identificar casos semelhantes de erros da avaliação.

6.1. Procedimentos para identificação de estudos

A expressão genérica de busca, definida a partir da amplitude da descrição de pesquisa, permite a correta identificação dos estudos provavelmente relevantes à revisão sistemática. No entanto, seria desejável projetar uma forma para assegurar que os artigos relevantes estão sendo realmente recuperados.

Uma forma de conseguir isso é pela análise da sensibilidade e a precisão da expressão de busca em comparação a um conjunto de estudos de controle (previamente selecionados). A sensibilidade representa a habilidade para identificar todos os estudos relevantes. Idealmente, a expressão deve retornar todos os estudos relevantes do grupo de controle. No entanto, uma expressão ampla, além de recuperar todos os estudos, tende a recuperar uma quantidade de estudos elevadas, de forma que a cardinalidade do grupo de controle é muito inferior àquela do grupo obtido. A precisão mede exatamente isto: a habilidade da expressão de busca em não retornar estudos irrelevantes (DIESTE; PADUA, 2007).

A expressão ideal é aquela com alta precisão e sensibilidade. No entanto, as duas propriedades geralmente são inversamente proporcionais, ou seja, a alta precisão acarreta em uma baixa sensibilidade e a alta sensibilidade geralmente resulta em uma baixa precisão (DIESTE; PADUA, 2007).

Para revisões sistemáticas, os valores recomendados para precisão e sensibilidade, com base nos resultados finais da revisão sistemática, são:

- Valores ótimos: sensibilidade superior a 0,80 e precisão entre 0,20 e 0,25.
- Valores aceitáveis: sensibilidade entre 0,72 e 0,80 e precisão entre 0,15 e 0,25.

Caso adotado um processo iterativo para a revisão sistemática como um todo, futuras interações podem utilizar os valores desses atributos em iterações anteriores para refinar a expressão de busca. No entanto, se adotado simplesmente um processo iterativo no contexto de cada fase da revisão sistemática, a utilização desses valores não é possível: como os resultados da revisão não foram obtidos, não é possível calcular esses valores. Nesse caso, a alternativa é o estabelecimento de um grupo de controle e utilizá-lo, durante o processo de seleção, para ajustar a expressão de busca. No entanto, não existem valores estabelecidos para este caso. Sugere-se que a precisão seja superior a 0,95 (dado que o grupo de controle já é reduzido) e que a sensibilidade seja definida em função da quantidade de artigos com precisão de 0,80 e com a precisão 0,95: se a quantidade de artigos recuperados com precisão aumentada for superior em XX% ou menos quanto com precisão reduzida, a sensibilidade está adequada; caso contrário, deve-se buscar alterações na expressão de busca para aumentar a sensibilidade, sem prejudicar a precisão.

O ajuste da expressão de busca é realizada principalmente pela alteração dos termos, mais especificamente dos sinônimos associados a cada palavra-chave definida originalmente. Eventualmente, faz-se necessária a utilização de um sinônimo ao invés da palavra-chave original. Neste caso, recomenda-se a revisão dos atributos de amplitude e verificar se é o caso do sinônimo ser realmente o termo mais apropriado. Como este ajuste pode ser realizado durante o ensaio do protocolo, as ameaças à validade são poucas (restritas ao viés introduzido pelos artigos utilizados como grupo de controle).

6.2. Procedimentos para a seleção de estudos

As palavras-chaves devem constar no documento (no título, no resumo ou nas palavras-chaves do resumo). O resumo dos estudos serão lidos e, se de acordo com os critérios de inclusão estabelecidos, serão selecionados.

Cada par analisará metade dos estudos. Cada par ainda analisará um subconjunto aleatório de estudos analisados por outros pares. Se os resultados obtidos não coincidirem, deve-se realizar uma reunião para a

identificação do conflito e revisão das análises efetuadas.

Os resultados da seleção devem ser anotados, preferencialmente no próprio registro bibliográfico do estudo. As seguintes informações devem ser registradas:

- Fonte da qual o presente estudo foi obtido. Como toda expressão de busca é unicamente identificada, esse identificar deve ser utilizado para definir este atributo. Observa-se que é possível que um mesmo estudo tenha sido obtido de diferentes fontes ou expressões de busca.
- Estado do artigo (indefinido, selecionado ou excluído).
- Justificativa para o estado atual do artigo.
- Estudo ao qual este presente artigo pertence. Dado que uma mesma pesquisa pode resultar em vários estudos, todos descrevendo o mesmo experimento, é necessário criar um meta-estudo para agrupar todos os estudos diretamente relacionados.

Recomenda-se a utilização da ferramenta JabRef (ALVER; OTHER, 2003) para a condução do processo de seleção e exclusão. Ela permite a definição de campos personalizados, pelas opções `Options` e `Set up general fields`. Para os dados específicos de seleção e exclusão de estudos, a seguinte configuração é sugerida:

```
Review : source ; status ; reasoning ; study_id
```

6.3. Arbitragem da lista e critérios

6.3.1. Seleção de estudos

6.3.2. Procedimentos para seleção de estudos

Os estudos recuperados devem ser cadastrados e suas informações (dados bibliográficos) devidamente armazenadas. Cada estudo deve ser unicamente identificado, de forma que seja possível sua posterior análise e identificação. Para tanto, uma solução é atribuir um número único, automaticamente incrementado, como identificador de cada artigo. Os dados bibliográficos podem ser recuperados no formato BibTeX (e a sua chave definida de acordo com o identificador previamente definido).

A maioria dos dispositivos de busca suporta a exportação de informações bibliográficas no formato BibTeX, com graus menores ou maiores de facilidade (*e.g.*, o Scopus permite a exportação de todos os estudos encontrados em um único passo enquanto a ACM DL requer o uso de ferramentas externas). As instruções para cada mecanismo de busca serão definidas nas subseções desta.

As ferramentas recomendadas para esta fase do processo de revisão sistemática são:

- **Firefox** (Mozilla Foundation, 2004). Todos os mecanismos de busca descritos neste relatório são aplicações Web, requerendo um navegador. Como o Firefox permite a utilização de extensões (como o Zotero, conforme descrito a seguir), ele é o navegador Web recomendado.
- **Internet Explorer** (Microsoft Corporation, 1995). Embora o Firefox seja o navegador recomendado, alguns sítios de busca não são totalmente compatíveis com ele (*e.g.*, o dispositivo para obter os textos completos dos estudos recuperados do Scopus funciona apenas com o Internet Explorer). Portanto,

recomenda-se a utilização do Internet Explorer para os sítios que apresentem incompatibilidades com o Firefox.

- **Zotero** (Center for History and New Media (George Mason University), 2006). Analisador e organizador de referências. Ele é uma extensão para o navegador Web Firefox (Mozilla Foundation, 2004) e permite a extração dos dados bibliográficos de materiais apresentados em uma página Web (particularmente útil para recuperar estudos encontrados por mecanismos de busca como a ACM).
- **JabRef** (ALVER; OTHER, 2003). Gerenciador de referências (em especial BibTeX). Embora o Zotero permita o gerenciamento de referências bibliográficas, o volume de estudos recuperados em revisões sistemática, bem como outros dados a serem acrescentados à essas referências durante a execução do estudo experimental, requerem a utilização de uma ferramenta executada localmente e que permita customizações. O JabRef possui um bom desempenho (mesmo com bases de milhares de referências) e permite a criação de novos campos de dados para cada referência (o que pode ser utilizado para controlar a execução do processo de revisão sistemática).

Como destacado anteriormente (Capítulo 5), as fontes de estudo selecionadas podem ser indexadas por um ou mais mecanismos de busca. Por exemplo, a *Digital Library*, da ACM, indexa publicações da ACM, IEEE, Springer-Verlag, dentre outros¹. Dessa forma, é necessário estabelecer mecanismos para remover estudos duplicados (e, ao mesmo tempo, manter a referência ao mecanismos de busca do qual ele foi recuperado). Uma solução simples é a adição de dois novos campos para descrever a fonte de dados e os mecanismos de busca das quais o estudo foi obtido, respectivamente **source** e **searchengine**. A partir dos resultados de cada operação, é trivial identificar o **searchengine**. Já para a identificação do **source**, é necessário analisar a referência. No entanto, este último campo não é tão relevante: sua principal utilidade é para verificar, posteriormente, se os mecanismos de busca estão recuperando corretamente os documentos. Sugere-se a identificação dos itens primeiramente pelo título e, posteriormente, pelos autores, local de publicação, ano e demais dados. Caso seja identificado que o estudo está duplicado, deve-se escolher uma das entradas como a principal e completar os seus dados com aqueles do outro estudo (caso existe algum dado adicional nele). Para o caso do campo **searchengine**, deve-se acrescentar o valor do campo do registro duplicado. Depois, deve-se apagar o estudo anterior.

6.4. Critérios de seleção baseados em VTM

Mineração visual de dados (VTD – *Visual Text Datamining* (VTD)) é um passo no processo de extração de conhecimento (KDD) que utiliza a visualização como um canal de comunicação entre computador e usuário para apoiar a identificação de padrões novos e interpretáveis (ANKERST, 2000).

As principais atividades desenvolvidas na VTD são: recuperação de dados, pré-processamento, formatação mineração e avaliação de padrões. No contexto específico de revisões sistemáticas, em que os dados são majoritariamente documentos, emprega-se a mineração visual de texto.

¹ A lista completa está disponível em http://portal.acm.org/contents_guide.cfm.

Extração

Selecionados os estudos, procede-se à extração dos dados relevantes deles. O protocolo da revisão sistemática estabelece os critérios para a obtenção dos dados e os formulários necessários. São dois os tipos de dados a serem obtidos: os objetivos e os subjetivos: os objetivos são extraídos diretamente dos estudos; os subjetivos são obtidos da análise crítica dos estudos pelo revisor e pela consulta aos autores.

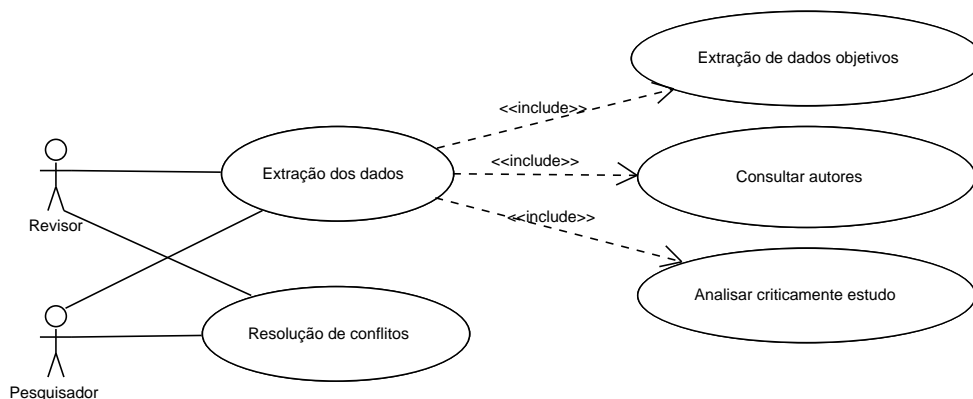


Figura 7.1: Caso de uso de extração de dados de estudos.

Da mesma maneira com que, na seleção dos estudos, sugere-se o emprego de um pesquisador para revisar os estudos selecionados, na extração também é recomendável que um pesquisador extraia os dados de um subconjunto, aleatoriamente determinado, de estudos e que os resultados de ambos sejam comparados. Caso o resultado seja distinto, deve-se resolver o conflito da extração e, caso confirmado o erro do protocolo, o mesmo deve ser revisado de modo a identificar o ponto que deve ser esclarecido para que conflitos semelhantes não ocorram nas próximas iterações. Caso seja confirmado um erro do revisor, deve-se revisar os demais dados extraídos e identificar casos semelhantes de erros da extração.

7.1. Extração de dados básicos dos artigos

Para cada artigo selecionado, os seguintes dados deverão ser extraídos:

- Artigo
 - Nome do artigo.
 - Fonte do artigo.
 - Ano do artigo.
 - Autores do artigo.
- Padrão
 - Nome do padrão.
 - Data de criação do padrão (ano e, opcionalmente, mês).
 - Autores do padrão (pessoa ou grupo de pesquisa).
 - Domínio de aplicação do padrão.
 - Escopo do padrão (produto ou processo).
 - Principais características do padrão (opcional).
 - Limitações do padrão (opcional).

Além dos dados mencionados, deverão ser obtido os demais necessários para especificar um item BibTeX.

Para artigos publicados em conferências, os dados são:

- Identificador (chave).
- Autores (`author`).
- Título (`title`).
- Conferência (`booktitle`).
- Local (`address`).
- Ano (`year`).
- Páginas (`pages`).
- Publicador (`publisher`).

Para artigos publicados em revistas, os dados são:

- Identificador (chave).
- Autores (`author`).
- Título (`title`).
- Jornal (`journal`).
- Ano (`year`).
- Volume (`volume`).
- Número (`number`).
- Páginas (`pages`).

Para os casos omissos de artigos, deverá ser anotado o endereço Web (URL) do documento (`url`). Importante ressaltar que todos os artigos selecionados, uma cópia deverá ser mantida no computador (de fato, para todo arquivo identificado – e não apenas selecionado – é desejável que exista uma cópia local).

7.2. Extração de evidências

Extraídos os dados básicos sobre os artigos, procede-se à extração das evidências, conforme descrito no protocolo, com base nas questões secundários estabelecidas para a revisão sistemática.

Análise

Extraídos os dados necessários dos estudos primários, deve-se aplicar um método de síntese (como, por exemplo, meta-análise) para obter as conclusões gerais a partir da análise dos estudos individuais. A execução desta fase segue o protocolo e depende diretamente o método de síntese escolhido. Os resultados dessa síntese devem ser apresentados em formato adequado (tabelas, gráficos).

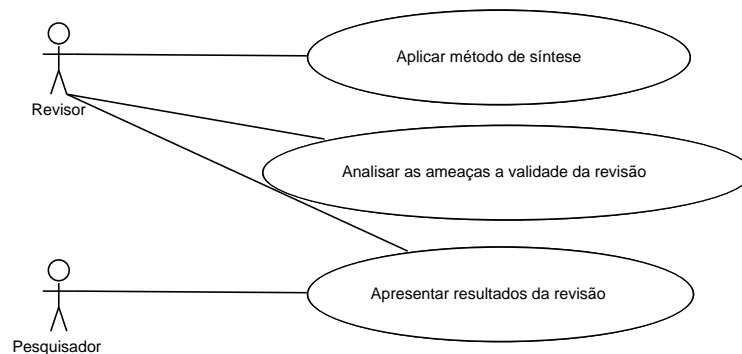


Figura 8.1: Caso de uso de análise de dados de estudos.

Além dos resultados específicos dos método de síntese, é necessário informar dados gerais da revisão sistemática: quantidade de estudos analisados, ameaças à validade da revisão sistemática, conflitos entre os atores no processo, sugestões de como aplicar os resultados e recomendações para possíveis replicadores da revisão sistemática.

O método de análise a ser utilizado depende de vários fatores quanto aos estudos em questão:

- Tipo de estudo experimental.
- Homogeneidade (ou heterogeneidade) dos dados.
- Tipos de dados (se não comparáveis).

- Dependência quanto ao valor do tamanho do efeito.

8.1. Contagem de voto

A contagem de votos é um dos métodos mais simples e fracos de síntese. Ela consiste na contagem de votos de cada estudo quanto ao foco da questão da revisão sistemática (BIOLCHINI *et al.*, 2005).

Considere o exemplo de uma revisão sistemática sobre a efetividade da programação pareada. A revisão sistemática encontrou e selecionou 35 estudos. Desses, 25 afirmam que a programação pareada é benéfica à produtividade e à qualidade do software produzido, 5 afirmam que não traz benefício algum e as cinco restantes são inconclusivas. Logo, seriam 25 votos a favor, 5 contra e cinco em branco.

8.2. Meta-análise

A meta-análise consiste na análise estatística de uma coleção de resultados de estudos individuais com o propósito de integrar os resultados citeglass:1976, como se eles fossem parte de um estudo maior. Em outras palavras, a síntese consiste na consideração que todos os estudos encontrados são, na verdade, um único estudo. Assim, é necessário unir os resultados dos trabalhos individuais em um único resultado final (BIOLCHINI *et al.*, 2005).

Embora a meta-análise seja de difícil execução (dado que os dados extraídos de cada estudo podem não ser compatíveis entre si ou a qualidade dos estudos é heterogênea), ela permite a obtenção de conclusões que não seriam possíveis pela observação e análise dos estudos individualmente (BIOLCHINI *et al.*, 2005).

8.3. Síntese meta-etnográfica

A síntese meta-etnográfica é adequada para a análise de dados qualitativos. Seu produto final é a tradução dos estudos entre eles, ou seja, como eles se relacionam entre si, considerando-se seu conteúdo. São três as formas possíveis de relacionamento (BRITTEN *et al.*, 2002):

- os estudos podem ser diretamente comparados como traduções recíprocas,
- os estudos podem estar em oposição um quanto ao outro (como traduções refutadoras),
- os estudos podem representar, juntos, uma linha de argumento.

A síntese meta-etnográfica organiza-se em sete passos (BRITTEN *et al.*, 2002):

1. Preparativos para a síntese.
2. Decisão sobre o que é relevante quanto ao interesse inicial do estudo (a revisão sistemática, no caso).
3. Leitura dos estudos.
4. Determinação de como os estudos estão relacionados.
5. Tradução dos estudos entre si.
6. Síntese das traduções.
7. Expressão da síntese.

Durante a leitura, realiza-se a identificação dos principais conceitos de cada estudo selecionado, tal como originalmente descrito no texto original (posteriormente os termos serão traduzidos).

O processo de tradução dos estudos e a síntese permite a identificação de temas ou conjuntos de interpretações de alto nível que são recorrentes entre os estudos, as dimensões positivas e negativas dos

resultados relatados e a identificação de lacunas dentre os temas identificados (DYBA *et al.*, 2007).

Empacotamento

A revisão sistemática deve ser documentada de modo a permitir a sua análise pelos pares e a sua replicação. Para tanto, é necessário que seja especificado o protocolo adotado na revisão sistemática e relatados os estudos selecionados, as evidências coletadas, a síntese das evidências, as conclusões e as ameaças à validade da revisão sistemática, tal como representado na Figura 9.1.

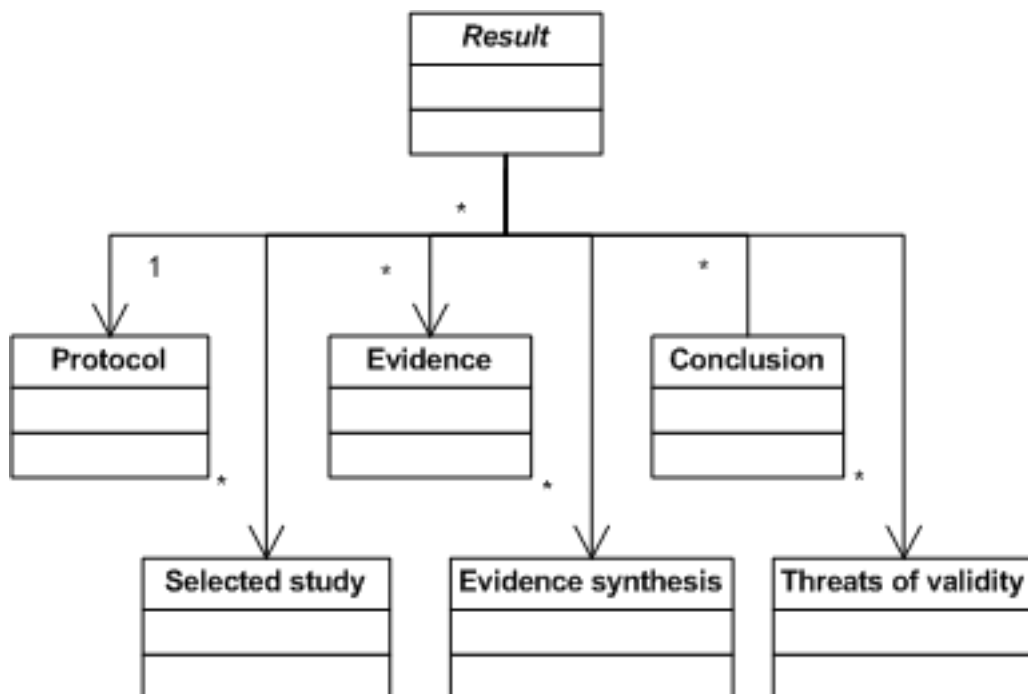


Figura 9.1: Resultados de uma revisão sistemática.

No caso de um processo de revisão sistemática iterativo, embora o protocolo permaneça essencialmente

o mesmo, a escolha de uma janela de tempo distinta da revisão inicial possibilita a obtenção de estudos, evidências, sínteses, conclusões e ameaças a validades distintas das iterações anteriores:

- Estudos novos serão obtidos em consequência da publicação de novos artigos e pela indexação de outros pré-existentes (na maioria dos mecanismos de recuperação de trabalhos científicos, a data de indexação é distinta daquela de publicação).

Um fato pouco comum, mas possível, é a retirada de um artigo de uma fonte de estudos. Isso pode ser decorrente da identificação de plágios.

Independentemente desses fatores, espera-se que o conjunto de estudos identificados e selecionados contenha parte dos estudos identificados em iterações anteriores.

- Uma consequência natural da seleção de novos estudos e da exclusão de estudos inválidos (*e.g.*, plágios), considerados anteriormente, é a alteração das evidências obtidas. Tal como no caso dos estudos selecionados, o conjunto de evidências obtidas contém um subconjunto das evidências obtidas nas iterações anteriores.
- Em virtude da alteração das evidências, é possível que a síntese das evidências seja distintas. O quão alterada será a análise depende da quantidade de evidências alteradas e da técnica de síntese. Não seria de se esperar que a síntese sofresse alterações muito significativas em virtude de novas iterações.
- As conclusões sofrem um grau de alteração proporcional àquele proporcionado à síntese de evidências.
- As ameaças à validade contém um subconjunto das ameaças identificadas anteriormente. Em relação aos elementos do protocolo e das técnicas de síntese de evidência, não se esperam alterações. No entanto, devido à natureza dos mecanismos de busca (que sofrem alterações), novas ameaças podem ser caracterizadas quanto à recuperação de estudos.

Embora o modelo sugira que existe apenas um protocolo para diversas revisões sistemáticas, a alteração do protocolo, em resposta às ameaças identificadas em iterações anteriores, é passível de realização. Nesse caso, deve-se documentar as alterações e analisá-las meticulosamente, garantindo-se que os resultados obtidos serão mais confiáveis do que aqueles apresentados anteriormente. Um exercício seria a aplicação das alterações nas iterações realizadas até então, considerando os estudos recuperados.

Em todo os casos, observa-se a importância de realizar o controle de configuração da revisão sistemática, armazenando-se todos os dados recuperados. inclusive os dados sobre os estudos recuperados (e não apenas daqueles selecionados). Dessa forma, pode-se repetir a revisão (desconsiderando-se a recuperação) e analisar o viés inserido pelo protocolo e pela equipe que conduz a revisão.

Conclusões

O objetivo da revisão sistemática é reunir as evidências de pesquisas primárias quanto a um tópico. Ela é um importante instrumento de uma área denominada Engenharia de Software Baseada em Evidências, por sua vez inspirada na Medicina Baseada em Evidências. A tese defendida é que o julgamento crítico do especialista é pior do que as conclusões obtidas a partir da análise sistemática das evidências. Não que o especialista seja dispensável, muito pelo contrário: o objetivo das engenharias baseadas em evidência é permitir que o especialista, com os resultados das pesquisas primárias – as evidências – obtenha, em suas pesquisas, resultados confiáveis e verificáveis pelos seus pares.

Uma revisão sistemática de qualidade é aquela que pode ser replicada por outros grupos. Essa replicação só é possível quando o protocolo é corretamente documentado e executado de acordo no estudo original. Desvios do mesmo provavelmente resultarão em conclusões que conflitarão com as das replicações. Poder-se-ia dizer então que uma revisão sistemática, *per si*, não é uma garantia de boas pesquisas e resultados até que ela seja replicada. De fato, isto é verdade. O protocolo permite a verificação da metodologia, mas não da execução. Assim também o é em qualquer experimento de engenharia de software. No entanto, acredita-se que as revisões (e demais técnicas da engenharia de software baseada em evidências) sejam mais facilmente replicáveis que os demais instrumentos utilizados para a validação de pesquisas de engenharia de software, permitindo, assim, um consistente avanço da pesquisa da área.

10.1. Ameaças

10.1.1. Protocolo

Duas importantes ameaças à validade de experimentos de engenharia de software são a dependência das habilidades dos participantes (engenheiros) e o ciclo de vida de engenharia de software.

A dependência quanto as habilidades dos engenheiros é uma questão importante porque ela impede o *blinding*, ou seja, a independência dos resultados do experimento quanto às pessoas envolvidas no experimento.

Na impossibilidade de garantir o *blinding* por projeto do experimento como um todo, ainda é possível minimizar os riscos reduzindo a interação direta entre os experimentadores e os objetos de estudo. As seguintes estratégias podem ser utilizadas:

- Alocação as cegas do grupo de tratamento (alocação aleatória de sujeitos para cada técnica).
- Distribuição as cegas (aleatória) da distribuição de material.
- Marcação as cegas ou automatizada. Se os resultados de uma tarefa não pode ser diretamente relacionado ao tratamento, os marcadores não devem saber qual tratamento foi utilizado pelos sujeitos.
- Análise as cegas. Os resultados devem ser codificados de modo que o analista não saiba o tratamento que será utilizado.
- Coleção as cegas dos dados. Sistemas computadorizados podem ser utilizados para armazenar os dados, garantindo a anonimidade no acesso aos dados.

O ciclo de vida de desenvolvimento de software, com a interação de várias técnicas e procedimentos em um mesmo projeto de software, dificulta o estabelecimento da relação causal entre o elemento sob estudo (por exemplo, uma técnica particular) e os resultados esperados. Duas alternativas podem ser utilizadas para reduzir esse problema: experimentar as técnicas individualmente (independentemente das demais técnicas) e conduzir estudos empíricos de larga escala. Infelizmente, o primeiro nem sempre é possível e o segundo possui a questão de como executá-lo com o rigor requerido. Uma alternativa é a realização, ao invés de experimentos, de quasi-experimentos e sintetizar os dados a partir de uma amostra aleatória de projetos.

10.1.2. Qualidade do experimento

Avaliação da qualidade das evidências utilizadas deve ser avaliada de acordo com os seguintes critérios:

- **Força (*strength*):** A força é caracterizada por três elementos: nível, qualidade e precisão estatística. O nível está relacionado à escolha do projeto de estudo e é utilizado como indicador de como o viés foi eliminado do desenho do experimento. Qualidade refere-se ao método de investigação utilizado para reduzir o viés. Precisão estatística é o valor P ou intervalo de confiança.
- **Tamanho do efeito:** Distância do efeito estimado do tratamento para o valor nulo e a inclusão de efeitos importantes no intervalo de confiança.
- **Relevância:** Utilidade da evidência na prática, particularmente o quão apropriados são os resultados esperados medidos.

10.1.3. Replicação

A replicação de experimentos deve ser encorajada. A detalhada e completa especificação da metodologia empregada é uma forma de incentivar a replicação. No entanto, cuidado deve ser tomado para que as replicações não sejam idênticas. As variações no projeto experimental e materiais utilizados reduzem o risco de introdução de viés nas replicações.

10.2. Documentação

10.3. Garantia de qualidade

Os estudos experimentais trazem, à Engenharia de Software, um rigor ausente e necessário às pesquisas atuais. No entanto, é necessário cautela ao afirmar que as evidências obtidas são suficientes para a tomada de qualquer decisão no processo de engenharia de software. De fato, o objetivo da engenharia de software experimental é prover os meios para que as melhores evidências científicas sejam combinadas com experiências práticas e valores humanos no processo de decisão relacionado ao desenvolvimento de software (KITCHENHAM *et al.*, 2004).

Por melhor que seja a execução dos estudos experimentais, sempre existem riscos inerentes à sua execução e resultados. Por este motivo é que as ameaças à validade de todos eles devem ser claramente identificadas e relatadas. De fato, quanto maior o rigor do estudo, maior a confiança atribuída às evidências obtidas – mas a confiança nunca alcança o valor de 100%. Assim, mesmo para o caso real “ótimo”, de um processo de desenvolvimento com exatamente as mesmas condições relatadas para um experimento, é necessário o julgamento humano para decidir se realmente o resultado é aplicável. No entanto, é certo que a decisão tomada será fundamentada, de modo que a probabilidade da escolha realizada resultar em um fracasso, é reduzida e, ainda assim, nesse caso, será mais fácil identificar a razão deste fracasso, o que reconduzirá, em uma situação futura, a uma decisão com probabilidade maior de sucesso.

10.4. Trabalhos futuros

A associação de Experimental com Engenharia, proposta no nome da disciplina, Engenharia de Software Experimental, soa estranha. Afinal, é uma disciplina de Engenharia, então o método de pesquisa não deveria ser o de engenharia? Aqui pode-se formular duas hipóteses: (1) o método experimental é o mais apropriado para Engenharia ou (2) Engenharia de Software não é Engenharia. Quanto ao último argumento, existem aqueles que a defendem (DENNING; RIEHLE, 2009), dada as características únicas da área. Quanto à primeira, intuitivamente percebe-se que o método experimental talvez não seja o mais apropriado, mas certamente é um importante método. De fato, a utilização do método experimental não exclui o uso do método de engenharia. Talvez seja a questão de que a Engenharia de Software está tão carente de resultados científicos rigorosos que o método experimental é o único que permite a obtenção de resultados de forma unívoca.

Os estudos secundários, que são uma forma de pesquisa do método experimental, permitem observar a relação das evidências com outras pesquisas. A identificação, o acompanhamento das pesquisas que usam, como argumentos, as evidências obtidas em revisões sistemáticas, e a identificação dos métodos de pesquisa desses estudos, permitiria observar a tendência quanto a utilização de outros métodos de pesquisa e o rigor quanto a utilização desses métodos. Espera-se que pesquisas de qualidade com o método experimental tenham, como consequência, a condução de pesquisas de qualidade com outros métodos.

Referências

ALVER, Morten O.; OTHER. *JabRef*. 2003. Programa de computador. Disponível em: <http://jabref.sourceforge.net/>.

ANKERST, M. *Visual Data Mining*. Tese (Doutorado) — Faculty of Mathematics and Computer Science, University of Munich, 2000.

BABBIE, Earl R. *Survey Research Methods*. 2. ed. [S.l.]: Wadsworth, 1990. 395 p.

BALDASSARRE, Maria Teresa; BOFFOLI, Nicola; CAIVANO, Danilo; VISAGGIO, Giuseppe. A hands-on approach for teaching systematic review. In: *International Conference on Product-Focused Software Process Improvement*. Monte Porzio Catone, Italy: Springer-Verlag, 2008. (Lecture Notes in Computer Science, 5089), p. 415–426. ISBN 978-3-540-69564-6.

BASILI, Victor R. The experimental paradigm in software engineering. In: *International Workshop on Experimental Software Engineering Issues: Critical Assessment and Future Directions*. London, UK: Springer-Verlag, 1992. (Lecture Notes in Computer Science), p. 3–12.

BASILI, V. R.; SHULL, F.; LANUBILE, F. Building knowledge through families of experiments. *IEEE Transactions of Software Engineering*, v. 25, n. 4, p. 456–474, jul. 1999.

BERGIN, Shan; WRAIGHT, Paul. Silver based wound dressings and topical agents for treating diabetic foot ulcers. *Cochrane Database of Systematic Reviews*, 2006.

BIOLCHINI, Jorge; MIAN, Paula Gomes; NATALI, Ana Candida Cruz; TRAVASSOS, Guilherme Horta. *Systematic Review in Software Engineering*. Rio de Janeiro, RJ, Brazil, maio 2005. Disponível em: alarcos.inf-cr.uclm.es/doc/MetoTecInfInf/Articulos/es67905.pdf.

BOEHM, Barry; ROMBACH, Hans Dieter; ZELKOWITZ, Marvin V. (Ed.). *Foundations of Empirical Software Engineering: The Legacy of Victor R. Basili*. Germany: Springer-Verlag, 2005. 431 p.

BRERETON, Pearl; KITCHENHAM, Barbara A.; BUDGEN, David; TURNER, Mark; KHALIL, Mohamed. Lessons from applying the systematic literature review process within the software engineering domain. *The Journal of Systems and Software*, v. 80, n. 80, p. 571–583, 2007.

BRIAND, L. C.; EMAM, K. El.; MORASCA, S. On the application of measurement theory in software engineering. *Journal of Empirical Software Engineering*, v. 1, n. 1, p. 61–88, 1996.

BRITTEN, Nicky; CAMPBELL, Rona; POPE, Catherine; DONOVAN, Jenny; MORGAN, Myfanwy; PILL, Roisin. Using meta ethnography to synthesise qualitative research: a worked example. *Journal of Health Services Research & Policy*, v. 7, n. 4, p. 209–215, out. 2002. Disponível em: <http://jhsrp.rsmjournals.com/cgi/content/abstract/7/4/209>.

BUDGEN, David; KITCHENHAM, Barbara A.; CHARTERS, Stuart M.; TURNER, Mark; BRERETON, Pearl; LINKMAN, Stephen G. Presenting software engineering results using structured abstracts: a randomised experiment. *Empirical Software Engineering*, Kluwer Academic Publishers, Hingham, MA, USA, v. 13, n. 4, p. 435–468, ago. 2008. ISSN 1382-3256. Disponível em: <http://portal.acm.org/citation.cfm?id=1409479.1409483>.

Center for History and New Media (George Mason University). *Zotero*. out. 2006. Programa de computador. Disponível em: <http://www.zotero.org/>.

COOK, T.; CAMPBELL, D. *Quasi-Experimentation – Design and Analysis Issues for Field Settings*. [S.l.]: Houghton Mifflin Company, 1979.

DENNING, Peter J.; RIEHLE, Richard D. The profession of it: Is software engineering engineering? *Communications of the ACM*, ACM, New York, NY, USA, v. 52, n. 3, p. 24–26, mar. 2009. ISSN 0001-0782.

DIESTE, O.; PADUA, O.A.G. Developing search strategies for detecting relevant experiments for systematic reviews. In: *International Symposium on Empirical Software Engineering and Measurement*. Madrid, Spain: [s.n.], 2007. p. 215 –224. ISSN 1938-6451.

DIXON-WOODS, M.; AGARWAL, S.; JONES, D.; YOUNG, B.; SUTTON, A. Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Services Research & Policy*, Royal Society of Medicine Press, v. 10, n. 1, p. 45–53, jan. 2005.

DYBA, T.; DINGSOYR, T.; HANSSEN, G.K. Applying systematic reviews to diverse study types: An experience report. In: *International Symposium on Empirical Software Engineering and Measurement*. Washington, DC, EUA: IEEE Computer Society, 2007. p. 225 –234. ISSN 1938-6451.

DYBÄ, Tore; KAMPENES, Vigdis By; SJÄYBERG, Dag I.K. A systematic review of statistical power in software engineering experiments. *Information and Software Technology*, Elsevier, v. 48, n. 8, p. 745 – 755, ago. 2006. ISSN 0950-5849.

FENTON, N. Software measurement: A necessary scientific basis. *IEEE Transactions on Software Engineering*, v. 3, n. 20, p. 199–206, mar. 1994.

FENTON, Norman; PFLEEGER, Sharl Lawrence. *Software Metrics: A Rigorous & Practical Approach*. 2. ed. [S.l.]: Course Technology, 1998. 656 p.

FERRARI, Fabiano Cutigi; MALDONADO, JosÃ Carlos. Experimenting with a multi-iteration systematic review in software engineering. In: *Experimental Software Engineering Latin American Workshop (V ESELAW)*. Salvador, Brazil: [s.n.], 2008. p. 1–10.

GLASS, Robert L. The (solved, unsolved) problem of literature searches. *Journal of Systems and Software*, v. 15, p. 203–204, 1991.

GLASS, Robert L. The software research crisis. *IEEE Software*, v. 11, p. 42–47, nov. 1994.

GLASS, Robert L. An assessment of systems and software engineering scholars and institutions, 1993 and 1994. *Journal of Systems and Software*, v. 31, n. 1, p. 3 – 6, 1995. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-404RP1T-K/2/e882ce9aa69893440885f8a93b36dafc>.

GLASS, Robert L. An assessment of systems and software engineering scholars and institutions (1993-1995). *Journal of Systems and Software*, v. 35, n. 1, p. 85 – 89, 1996. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-3VTB04S-J/2/f425b15e95b9e1b458d84c4bbdc46467>.

GLASS, Robert L. An assessment of systems and software engineering scholars and institutions (1993-1997). *Journal of Systems and Software*, v. 43, n. 1, p. 59 – 64, 1998. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-3VN3C8N-7/2/32a6b1aac34ad34bec1c7faeabae809c>.

GLASS, Robert L. An assessment of systems and software engineering scholars and institutions (1994-1998). *Journal of Systems and Software*, v. 49, n. 1, p. 81 – 86, 1999. ISSN 0164-1212. See also Corrigendum to: An assessment of systems and software engineering scholars and institutions (1994-1998). Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-3Y9RCX5-8/2/7d2645007751ff2ab2ad9667da2c9fd6>.

GLASS, Robert L. An assessment of systems and software engineering scholars and institutions (1995-1999). *Journal of Systems and Software*, v. 54, n. 1, p. 77 – 82, 2000. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-419BHW-9/2/7ddef79e1fca9675719cfee65e691699>.

GLASS, Robert L. Corrigendum to: An assessment of systems and software engineering scholars and institutions (1994-1998). *Journal of Systems and Software*, v. 51, n. 3, p. 275 – 275, 2000. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-4007R6S-9/2/504d6a811b64671f1ce5ef0c0e2515fd>.

GLASS, Robert L.; CHEN, T.Y. An assessment of systems and software engineering scholars and institutions (1999-2003). *Journal of Systems and Software*, v. 76, n. 1, p. 91 – 97, 2005. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-4DCMP5X-1/2/83cd4d2592fef6678870a3451046e77a>.

GLASS, Robert L.; CHEN, T. Y. An assessment of systems and software engineering scholars and institutions (1996-2000). *Journal of Systems and Software*, v. 59, n. 1, p. 107 – 113, 2001. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-449THV4-9/2/2cfd3945ff1bcb5a7278f9ce91fa3d3d8>.

GLASS, R. L.; CHEN, T. Y. An assessment of systems and software engineering scholars and institutions (1997-2001). *Journal of Systems and Software*, v. 64, n. 1, p. 79 – 86, 2002. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-47T8NBW-1/2/62bdfc63328f61e66f443128035769c2>.

GLASS, R. L.; CHEN, T. Y. An assessment of systems and software engineering scholars and institutions (1998-2002). *Journal of Systems and Software*, v. 68, n. 1, p. 77 – 84, 2003. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-491RV6C-2/2/9b3c21b6ef4263e926f57fbfb6371459>.

IEEE. *IEEE Standard Glossary of Software Engineering Terminology*. set. 1990. Standard.

JURISTO, Natalia; MORENO, Ana M. *Basics of Software Engineering Experimentation*. [S.l.]: Kluwer Academic Publishers, 2001. 395 p.

KITCHENHAM, Barbara. *Procedures for Performing Systematic Reviews*. Keele, Reino Unido, jul. 2004. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.3308&rep=rep1&type=pdf>.

KITCHENHAM, B.A.; BRERETON, O.P.; OWEN, S.; BUTCHER, J.; JEFFERIES, C. Length and readability of structured software engineering abstracts. *IET Software*, v. 2, n. 1, p. 37–45, fev. 2008. ISSN 1751-8806.

KITCHENHAM, B.A.; PFLEEGER, S.L.; PICKARD, L.M.; JONES, P.W.; HOAGLIN, D.C.; EMAM, K. El; ROSENBERG, J. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on Software Engineering*, v. 28, n. 8, p. 721–734, ago. 2002. ISSN 0098-5589.

KITCHENHAM, B.; PICKARD, L.; PFLEEGER, S. L. Case studies for method and tool evaluation. *IEEE Software*, v. 12, p. 52–62, jul. 1995.

KITCHENHAM, Barbara A.; DYBA, Tore; JORGENSEN, Magne. Evidence-based software engineering. In: *International Conference on Software Engineering (ICSE'04)*. Edinburgh, UK: [s.n.], 2004. p. 273–281.

LANDIS, J. Richard; KOCH, Gary G. The measurement of observer agreement for categorical data. *Biometrics*, International Biometric Society, v. 33, n. 1, p. 159–174, mar. 1977. ISSN 0006341X. Disponível em: <http://www.jstor.org/stable/2529310>.

MAFRA, Sãmulo Nogueira; TRAVASSOS, Guilherme Horta. *Estudos PrimÃqrios e SecundÃqrios Apoiando a busca por EvidÃncias em Engenharia de Software*. Rio de Janeiro, RJ, Brazil, mar. 2006. Disponível em: www.cos.ufrj.br/uploadfiles/1149103120.pdf.

MIAN, Paula; CONTE, Tayana; NATALI, Ana; BIOLCHINI, Jorge; TRAVASSOS, Guilherme. A systematic review process for software engineering. In: *Software Engineering Latin American Workshop (ESELAW)*. Kaiserslautern, Germany: [s.n.], 2005. (Workshop Series on Empirical Software Engineering, v. 1), p. 1–6. Workshop help in Brazil.

Microsoft Corporation. *Internet Explorer*. ago. 1995. Programa de computador. Disponível em: <http://www.microsoft.com/windows/internet-explorer/>.

MILLER, James. Applying meta-analytical procedures to software engineering experiments. *Journal of Systems and Software*, Elsevier Science Inc., New York, NY, USA, v. 54, n. 1, p. 29–39, set. 2000. ISSN 0164-1212.

MONTGOMERY, Douglas C. *Design and Analysis of Experiments*. 5. ed. [S.l.]: Wiley, 2000. 672 p.

Mozilla Foundation. *Firefox*. 2004. Programa de computador. Disponível em: <http://www.mozilla.com/firefox/>.

MULROW, C.; COOK, D. *Systematic Reviews: Synthesis of Best Evidence for Health Care Decisions*. American College of Physicians, 1998. Disponível em: <http://books.google.com/books?id=rYWIojXyioIC&lpg=PA122&ots=anHG-1CqjG&dq=%22systematic%20reviews%3A%20synthesis%20of%20best%20evidence%20for%20health%20care%20decisions%22&pg=PA5#v=onepage&q&f=false>.

OATES, Briony J; CAPPER, Graham. Using systematic reviews and evidence-based software engineering with masters students. In: *International Conference on Evaluation and Assessment in Software Engineering (EASE)*. Durham University, UK: [s.n.], 2009. p. 1–9. Disponível em: <http://www.bcs.org/content/conWebDoc/25028>.

- PETTICREW, Mark; ROBERTS, Helen. *Systematic Reviews in the Social Sciences: A Practical Guide*. [S.l.]: Blackwell, 2006. 336 p.
- PFLEEGER, Shari Lawrence. Experimental design and analysis in software engineering: Types of experimental design. *SIGSOFT Software Engineering Notes*, ACM, New York, NY, USA, v. 20, n. 2, p. 14–16, abr. 1995. ISSN 0163-5948.
- SACKETT, D. L.; STRAUS, S. E.; RICHARDSON, W. S. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2. ed. [S.l.]: Churchill Livingstone, 2000.
- SACKS, Henry S.; BERRIER, Jayne; REITMAN, Dinah; ANCONA-BERK; CHALMERS, Thomas C. Meta-analyses of randomized controlled trials. *The New England Journal of Medicine*, v. 316, p. 450–455, fev. 1987.
- SANTOS, Cristina Mamãldio da Costa; PIMENTA, Cibele Andrucioli de Mattos; NOBRE, Moacyr Roberto Cuce. A estratãlgia pico para a construãgãčo da pergunta de pesquisa e busca de evidãncias. *Revista Latino-Americana de Enfermagem*, v. 15, n. 3, p. 508–511, maio-jun. 2007.
- SHULL, Forrest; FELDMANN, Raimund L. Guide to advanced empirical software engineering. In: _____. [S.l.]: Springer, 2008. cap. Building Theories from Multiple Evidence Sources, p. 337–364.
- SHULL, Forrest; SEAMAN, Carolyn; ZELKOWITZ, Marvin. Victor R. Basili’s contributions to software quality. *IEEE Software*, v. 23, n. 1, p. 16–18, jan. 2006.
- STAKE, Robert E. *The Art of Case Study Research*. 1. ed. [S.l.]: Sage Publications, 1995. 192 p.
- TRAVASSOS, Guilherme Horta; GUROV, Dmytro; AMARAL, Edgar Augusto Gurgel do. *Introduãgãčo ãã Engenharia de Software Experimental*. Rio de Janeiro, RJ, Brazil, 2002. Disponível em: <http://www.ufpa.br/cdesouza/teaching/topes/4-ES-Experimental.pdf>.
- TSE, T. H.; CHEN, T. Y.; GLASS, Robert L. An assessment of systems and software engineering scholars and institutions (2000-2004). *The Journal of System and Software*, v. 79, p. 816–819, 2004.
- WARREN, K. S.; MOSTELLER, F. Doing more good than harm – the evaluation of health care interventions. In: *New York Academy of Sciences*. [S.l.: s.n.], 1993.
- WOHLIN, Claes; RUNESON, Per; HãúST, Martin; OHLSSON, Magnus C.; REGNELL, Bjãúrn; WESSLãIN, Anders. *Experimentation in Software Engineering: An Introduction*. 1. ed. Sweden: Kluwer Academic Publishers, 2000. 204 p. (The Kluwer International Series in Software Engineering).
- WONG, W. Eric; TSE, T.H.; GLASS, Robert L.; BASILI, Victor R.; CHEN, T.Y. An assessment of systems and software engineering scholars and institutions (2001-2005). *Journal of Systems and Software*, v. 81, n. 6, p. 1059 – 1062, 2008. ISSN 0164-1212. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-4PRRBG9-1/2/568653a276de312b23ab5f1de33cac9a>.
- WONG, W. Eric; TSE, T.H.; GLASS, Robert L.; BASILI, Victor R.; CHEN, T.Y. An assessment of systems and software engineering scholars and institutions (2002-2006). *Journal of Systems and Software*, v. 82, n. 8, p. 1370 – 1373, 2009. ISSN 0164-1212. SI: Architectural Decisions and Rationale. Disponível em: <http://www.sciencedirect.com/science/article/B6V0N-4WJBBPB-1/2/d366961873cef17aca306dd20bbbabb2>.
- WONG, W. Eric; TSE, T.H.; GLASS, Robert L.; BASILI, Victor R.; CHEN, T.Y. An assessment of systems and software engineering scholars and institutions (2003-2007 and 2004-2008). *Journal of Systems and Software*, v. 84, n. 1, p. 162 – 168, 2011. ISSN 0164-1212. Information Networking and Software Services.

WOODALL, Phillip; BRERETON, Pearl. Conducting a systematic literature review from the perspective of a ph.d. student. In: *International Conference on Evaluation and Assessment in Software Engineering (EASE)*. [S.l.: s.n.], 2006.

YIN, Robert K. *Case Study Research: Design and Methods*. 3. ed. [S.l.]: Sage Publications, 2002. 200 p.

ZELKOWITZ, Marvin V. An update to experimental models for validating computer technology. *The Journal of Systems and Software*, v. 82, p. 373–376, 2008.

ZELKOWITZ, Marvin V.; WALLACE, Dolores R. Experimental models for validating computer technology. *IEEE Computer*, v. 31, n. 5, p. 23–31, maio 1998.