

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**CRITÉRIOS PARA PRIORIZAÇÃO DE ESTUDOS
PRIMÁRIOS IDENTIFICADOS POR SNOWBALLING
COM CONJUNTO INICIAL GERADO POR STRING DE
BUSCA**

CLEITON RODRIGO QUEIROZ SILVA

ORIENTADORA: PROFA. DRA. SANDRA FABBRI

São Carlos - SP
Janeiro/2017

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**CRITÉRIOS PARA PRIORIZAÇÃO DE ESTUDOS
PRIMÁRIOS IDENTIFICADOS POR SNOWBALLING
COM CONJUNTO INICIAL GERADO POR STRING DE
BUSCA**

CLEITON RODRIGO QUEIROZ SILVA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Engenharia de Software
Orientadora: Dra. Sandra Camargo Pinto Ferraz Fabbri.

São Carlos - SP
Janeiro/2017



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de Dissertação de Mestrado do candidato Cleiton Rodrigo Queiroz Silva, realizada em 16/02/2017.

Profª. Drª. Sandra Camargo Pinto Ferraz Fabbri
(UFSCar)

Prof. Dr. Fabiano Cutigi Ferrari
(UFSCar)

Profª. Drª. Katia Romero Felizardo
(USP)

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Katia Romero Felizardo, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do aluno Cleiton Rodrigo Queiroz Silva.

Profª. Drª. Sandra Camargo Pinto Ferraz Fabbri
Presidente da Comissão Examinadora
(UFSCar)

AGRADECIMENTOS

Agradeço a minha orientadora, a professora Sandra, por todos os ensinamentos e oportunidades que me foram oferecidos. Sou grato pela compreensão em momentos que precisei e por ter me auxiliado a melhorar sempre cada dia mais ao longo desses últimos anos em que convivemos e por sempre ser um exemplo a ser seguido. Agradeço por ter tido a oportunidade parte do grupo de pesquisa e dos projetos desenvolvidos no laboratório e sempre lembrarei com carinho dos momentos que me foram proporcionados.

Agradeço a meus familiares por todo o apoio e incentivo, agradeço a minha mãe Maria e meu pai Sebastião. Agradeço minha tia Zulmira e ao meu irmão Vinicius, por sempre estarem presentes quando precisei.

Agradeço a todos os amigos do LaPES, das novas gerações e das antigas, membros oficiais ou frequentadores assíduos. Agradeço a todos por toda a ajuda, sugestões, bate-papos durante o café e eventos e churrascos que compartilhamos. Um grande abraço a todos e em especial para Anderson Belgamo, Kamila Camargo, Juciara Nepomuceno, Guilherme, Arlindo, Augusto Zamboni, Rafael Gastaldi, Erik, Thiago Gaspar, Jesus, Odair, Ana Elisa, Luiz Cavamura, Matheus, Daise, Gabriel Malinpena, Elis Hernandez, Ivan e André Abade. Também ao Bento, Kathiani e ao Alessandro. Muito Obrigado.

Um agradecimento em especial para o Fábio, que me ajudou imensamente com os estudos, desenvolvimento da ferramenta e incontáveis revisões de texto. Muito Obrigado.

Agradeço ao professor André, por ter me guiado a ferramenta StArt e ao LaPES durante a iniciação científica. Agradeço por todo o incentivo e lições valiosas que me ensinou. Muito Obrigado.

Agradeço ao professor Fabiano e ao professor Auri, que sempre se mostraram dispostos a ajudar e me forneceram valiosas sugestões para a continuidade do meu trabalho.

Agradeço aos colegas e aos amigos de departamento pela convivência. Muito Obrigado a todos. Um abraço para especial para o Jãum, o Guido, o Victinho e o Diego.

RESUMO

Contexto: Em qualquer área da ciência, toda tecnologia proposta deveria ser avaliada e caracterizada para que a comunidade possa entender quando, onde e como aplicá-la. Particularmente, na área de Engenharia de Software, a preocupação constante com essa caracterização e a própria caracterização da tecnologia é referenciada como Engenharia de Software Baseada em Evidência (ESBE). Estudos secundários contribuem para construir essa caracterização porque utilizam métodos mais rigorosos de pesquisa. *Snowballing* é uma técnica de busca de estudos primários utilizada em estudos secundários. Ela é bem aceita pela comunidade, mas pode demandar um grande esforço em sua execução dependendo das diretrizes adotadas e do número de estudos primários relevantes existentes. **Objetivo:** Propor critérios de priorização de estudos primários para *snowballing* com conjunto inicial formado por buscas com *strings* para melhorar a eficiência na aplicação da técnica, fornecendo suporte computacional para sua aplicação. **Método:** Foram propostos quatro critérios de priorização de estudos primários e desenvolvidos algoritmos que permitem o suporte computacional à aplicação dos critérios. A avaliação foi realizada por meio de estudos experimentais, sendo um estudo exploratório e dois experimentos. **Resultados:** Os melhores resultados foram obtidos ao combinar os quatro critérios propostos, resultando em uma precisão de 96,46% e *recall* de 51,24%, tendo um percentual médio de erro de 20%. Os algoritmos criados para reconhecimento de referências e cálculo de similaridade apresentaram resultados satisfatórios. **Conclusão:** Com base nos resultados obtidos, há indícios de que os critérios de priorização propostos, que fazem uso dos algoritmos de extração de referências e de similaridade desenvolvidos e implementados na ferramenta StArt, contribuem para o aumento da eficiência na aplicação da técnica *snowballing* sem que haja perda significativa de qualidade dos resultados.

ABSTRACT

Background: In any area of science, all proposed technology should be evaluated and characterized so that the community can understand when, where and how to apply it. Particularly in the Software Engineering field, the concern with this characterization and the characterization of the technology is known as Evidence-Based Software Engineering (EBSE). Secondary studies contribute to this characterization because they use more rigorous research methods. Snowballing is a technique to identify relevant primary studies in secondary studies. It is well accepted by the community but may require a great deal of effort in its implementation depending on the guidelines adopted and on the number of existing relevant primary studies. **Objective:** To propose criteria for prioritizing primary studies for applying snowballing using the results obtained from search strings as seed aiming at improving the efficiency when applying the technique, providing computational support for its application. **Method:** Four criteria were proposed for prioritization of primary studies and algorithms were developed that allow computational support for applying the criteria. The evaluation was carried out by means of experimental studies, being an exploratory study and two experiments. **Results:** The best results were obtained by combining the four proposed criteria, resulting in a precision of 96.46% and a recall of 51.24%, with an average error rate of 20%. Algorithms created for reference extraction and similarity calculation presented satisfactory results. **Conclusion:** Based on the obtained results, there are indications that the proposed prioritization criteria, which make use of the reference extraction and similarity algorithms developed and implemented in the StArt tool, contribute to increase the efficiency of the snowballing technique without having a significant loss of quality of the results.

LISTA DE FIGURAS

Figura 1. Base de dados em que os estudos aceitos foram encontrados	38
Figura 2. Número de estudos por tipo de <i>snowballing</i>	39
Figura 3. Distribuição temporal do número de publicações encontradas	42
Figura 4. Exemplo de nuvem de citação (A), grafo de citação incompleto (B) e grafo de citação completo (C)	52
Figura 5. Exemplo de agrupamentos (<i>clusters</i>) no grafo de citação.	53
Figura 6 Grafo de citação completo gerado a partir do Quadro 1	59
Figura 7. Metáfora de técnicas de busca	61
Figura 8. Aba de referências extraídas do <i>full text</i> pela ferramenta StArt	76
Figura 9. Interface para a execução de <i>snowballing</i>	77
Figura 10. Interface da StArt para execução de <i>snowballing</i> manual com uso de PDF	78
Figura 11. Grafo de referências das referências na ferramenta StArt (Fabbri <i>et al.</i> , 2012)	78
Figura 12. Interface para visualizar agrupamento	79
Figura 13. Interface para visualização de nuvens de palavras.....	80
Figura 14. Avaliação de esforço dos participantes ao aplicar <i>snowballing</i>	90
Figura 15. Avaliação dos participantes sobre os recursos de snowballing na StArt	116
Figura 16 Plano de Trabalho da Ferramenta StArt.....	149
Figura 17 Catalogação de estudos na ferramenta StArt	150

LISTA DE TABELAS

Tabela 1. Relação do tipo de snowballing aplicado por área de conhecimento	39
Tabela 2. Relação de métodos de busca utilizados por área de conhecimento	40
Tabela 3. Técnicas para a escolha do conjunto inicial	63
Tabela 4. Relação de estratégias propostas para aplicação de <i>snowballing</i>	65
Tabela 5. Problemas relacionados aos critérios de priorização propostos.....	68
Tabela 6. Estratégias de paradas de <i>snowballing</i>	69
Tabela 7. Estratégias combinadas para <i>snowballing</i>	70
Tabela 8. Características das estratégias para a coleta de referências e citações ...	72
Tabela 9. Expressões regulares para reconhecimento do texto das referências.	74
Tabela 10. Descrição do experimento planejado	83
Tabela 11. Total de participantes por área de pesquisa.....	84
Tabela 12. Dados quantitativos referentes à aplicação de <i>snowballing</i> e dificuldades encontradas.....	89
Tabela 13. Dados quantitativos referentes à aceitação de estudos pelas técnicas...91	
Tabela 14. Porcentagem de estudos aceitos encontrados por meio de cada técnica	92
Tabela 15. Médias individuais de porcentagens de estudos identificados por método de busca.....	92
Tabela 16. Relação de referências encontradas e erro obtido	97
Tabela 17. Porcentagem de erro e acerto das estratégias.....	99
Tabela 18. Comparação entre os desempenhos dos algoritmos de comparação de referências.....	102
Tabela 19. Médias de erros e acertos na identificação das referências	103
Tabela 20. Médias de acertos e erros para os algoritmos de similaridade.....	103
Tabela 21. GQM planejado para o experimento.....	106
Tabela 22. Participantes por área de conhecimento	108
Tabela 23. Relação de resultados por grupo.....	112
Tabela 24. Dados das pesquisas dos participantes	114
Tabela 25. Porcentagem de estudos duplicados encontrados	115
Tabela 26. Total geral de referências e duplicações do estudo	116

Tabela 27. Relação de estudos encontrados por <i>snowballing</i> para estudos aceitos ou rejeitados.....	118
Tabela 28. Tabela de interpretação do Kappa	119
Tabela 29. Aplicação do coeficiente Kappa para os critérios propostos	120
Tabela 30. <i>Precision</i> e <i>recall</i> para os critérios aplicados na rejeição de estudos	122
Tabela 31. Critérios classificados pelo valor de <i>precision</i> (do maior para o menor)	124
Tabela 32. Critérios classificados pelos valores de <i>recall</i> (do maior para o menor)	124
Tabela 33. Efetividade dos critérios de priorização para rejeição	127
Tabela 34. Média dos valores de rejeição automática e erros para os critérios de priorização.....	128
Tabela 35. Protocolo do mapeamento sistemático.....	143
Tabela 36. Protocolo da revisão sistemática	146
Tabela 37. Conjunto de estudos aceitos por Hernandez, Belgamo e Fabbri (2013)	153

LISTA DE QUADROS

Quadro 1 Exemplo de iterações de <i>Snowballing</i> em grafos com conjunto inicial de um e dois elementos	58
---	----

LISTA DE ABREVIATURAS E SIGLAS

RS - Revisão Sistemática.

MS - Mapeamento Sistemático.

ESBE - Engenharia de Software Baseada em Evidência.

TI - Tecnologia da Informação.

MBE - Medicina Baseada em Evidência.

HTML - *HyperText Markup Language*.

API - *Application Programming Interface*

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	16
1.1 Contexto	16
1.2 Motivação e Objetivos	21
1.3 Metodologia de pesquisa.....	21
1.4 Organização do Trabalho	22
CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA.....	24
2.1 Considerações Iniciais.....	24
2.2 Estudos secundários em Engenharia de Software	25
2.3 Técnicas de busca para estudos secundários.....	27
• Busca com <i>Strings</i> (<i>String Search</i>)	28
• Bola de Neve (<i>Snowballing</i>)	29
• Busca por Literatura Cinzenta (<i>Gray Literature Search</i>)	29
• Opinião de Especialistas (<i>Expert Opinion</i>)	30
• Mapeamento de Grupos de Pesquisa (<i>Research Group Mapping</i>)	30
• Busca Manual (<i>Hand Search</i>)	31
2.4 Ferramentas de apoio para estudos secundários	32
2.5 Seleção semiautomática de estudos utilizando SCAS	33
2.6 Considerações Finais	34
CAPÍTULO 3 - LEVANTAMENTO BIBLIOGRÁFICO SOBRE <i>SNOWBALLING</i>	35
3.1 Considerações Iniciais.....	35
3.2 MS sobre o uso de <i>snowballing</i> em estudos secundários	36
• Uso de técnicas de busca em estudos secundários	41
3.3 RS sobre <i>snowballing</i> aplicado em estudos secundários.....	43
3.4 Considerações Finais	50
CAPÍTULO 4 - ESTRATÉGIAS PARA APLICAÇÃO DE <i>SNOWBALLING</i>	51
4.1 Subsídios para a definição de estratégias para aplicação de <i>snowballing</i> por meio buscas em grafos	51
4.2 Diretrizes para a execução de <i>snowballing</i> propostas por Wohlin (2014)	60

4.3	Técnicas para construção do conjunto inicial	61
4.3.1	Dificuldades de utilizar a busca com <i>strings</i> como <i>seed</i>	62
4.3.2	Estratégias propostas para composição do conjunto inicial	63
4.4	Crítérios para aplicar <i>snowballing</i> em estudos	64
4.4.1	Em quais estudos aplicar <i>snowballing</i> ?	64
4.4.2	Crítérios de Priorização de Referências	65
4.5	Quando encerrar a pesquisa utilizando <i>snowballing</i>	68
4.5.1	Estratégias para encerrar a pesquisa realizada com <i>snowballing</i>	69
4.6	Definindo uma estratégia completa para <i>snowballing</i>	69
4.7	Apoio computacional para <i>snowballing</i>	70
4.7.1	Extração de referências por meio do texto completo	70
4.7.2	Extração de referências por meio do arquivo de referências	71
4.7.3	Clientes de API Rest	71
4.7.4	Web Crawlers	72
4.7.5	Como identificar estudos em textos de referências e citações	73
4.7.6	Uso dos algoritmos de similaridade	74
4.7.7	Interface e Visualização	75
4.8	Considerações Finais	80
CAPÍTULO 5 - ESTUDOS EXPERIMENTAIS		81
5.1	Considerações Iniciais	81
5.2	Experimento: um comparativo entre <i>snowballing backward</i> e <i>forward</i> e as principais dificuldades de execução	82
5.2.1	Planejamento e Contexto do Experimento	82
5.2.2	Seleção das Variáveis	83
5.2.3	Seleção dos Participantes	84
5.2.4	Projeto Experimental	84
5.2.5	Instrumentação	85
5.2.6	Ameaças à Validade	85
5.2.7	Método	86
5.2.8	Coleta de Dados e Disponibilização dos dados	87
5.2.9	Análise e Interpretação	87
5.2.9.1	Análise Descritiva	88
5.2.9.2	Análise Estatística	91

5.3	Estudo exploratório: análise da eficácia das técnicas de extração de referências e dos algoritmos de similaridade ao aplicar <i>snowballing</i> com suporte da ferramenta StArt.	93
5.3.1	Seleção do Objeto de Estudo	93
5.3.2	Seleção das Variáveis.....	94
5.3.3	Método	94
5.3.4	Instrumentação	95
5.3.5	Questões a serem respondidas	95
5.3.6	Ameaças à Validade	95
5.3.7	Disponibilização dos Dados Coletados	96
5.3.8	Análise e Interpretação	96
5.3.8.1	Análise Descritiva.....	96
5.3.8.2	Análise Estatística.....	102
5.4	Experimento: a avaliação de critérios de priorização para aplicação de <i>snowballing backward</i> em busca com <i>String</i>	104
5.4.1	Planejamento e Contexto do Experimento.....	105
5.4.2	Seleção das Variáveis.....	106
5.4.3	Seleção dos Participantes.....	106
5.4.4	Projeto Experimental.....	107
5.4.5	Instrumentação	108
5.4.6	Questões a serem respondidas	108
5.4.7	Ameaças à Validade	109
5.4.8	Método	110
5.4.9	Coleta de Dados	111
5.4.10	Análise e Interpretação	111
5.4.10.1	Análise Descritiva.....	111
5.4.10.2	Análise Estatística.....	119
5.5	Considerações Finais.....	129
CAPÍTULO 6 - CONCLUSÃO		130
6.1	Conclusões	130
6.2	Contribuições e limitações da pesquisa	131
6.3	Lições aprendidas	133
6.4	Oportunidades Futuras	134

6.5	Publicações.....	135
•	Capítulos de livros publicados:	135
•	Trabalhos completos publicados em anais de congressos:	135
REFERÊNCIAS.....		137
APÊNDICE A - PROTOCOLO DO MAPEAMENTO SISTEMÁTICO		143
APÊNDICE B - PROTOCOLO DA REVISÃO SISTEMÁTICA.....		146
APÊNDICE C - A FERRAMENTA START.....		149
APÊNDICE D - REFERÊNCIAS DOS ESTUDOS DE CASO		153

Capítulo 1

INTRODUÇÃO

Este capítulo apresenta o contexto do trabalho caracterizando as dificuldades na aplicação da técnica snowballing, o que representa a principal motivação para propor melhorias em sua aplicação. Além disso, descreve-se a metodologia adotada na condução deste trabalho e apresenta-se como o texto está organizado.

1.1 Contexto

Kitchenham *et al.* (2004) definem o termo Engenharia de Software Baseada em Evidência (ESBE), que compreende métodos e técnicas para identificar evidências sobre tópicos de pesquisa na área de Engenharia de Software. Seu objetivo é auxiliar pesquisadores a obter evidências com maior qualidade na área. Segundo Jedlitschka e Ciolkowski (2004), uma evidência é a generalização de conclusões sobre um tópico específico partindo de um conjunto de estudos realizados, incluindo estudos experimentais. Evidências podem ser encontradas, por exemplo, na análise conjunta de estudos primários.

Segundo Mafra e Travassos (2006), estudos primários são estudos que visam a caracterizar uma tecnologia em uso dentro de um contexto específico. Greenhalgh

(2014) afirma que são pesquisas nas quais os autores têm envolvimento direto com os resultados e com as contribuições.

Dentre os métodos para a obtenção de evidências que compõem a ESBE, encontram-se os estudos secundários, que são realizados por meio de levantamentos bibliográficos feitos de maneira adequada. Estudos secundários são pesquisas que fazem uso de estudos primários e propõem evidenciar novos resultados com base na análise dos resultados de um conjunto de estudos primários (Maffra e Travassos, 2006). São estudos nos quais o pesquisador não tem envolvimento direto com o objeto de pesquisa estudado e sim executa uma análise com base nos resultados obtidos por outros pesquisadores. Os estudos secundários objetivam delinear as evidências contidas no conjunto de estudos primários coletados e relacionadas ao tópico de pesquisa em questão. Exemplos de estudos secundários são: revisão sistemática (RS) e mapeamento sistemático (MS).

Os estudos secundários têm por objetivo aumentar a qualidade do levantamento bibliográfico com a prática de uma metodologia sistemática por parte dos pesquisadores e, desde 2004, seu uso cresceu muito na área da Engenharia de Software, uma vez que já são empregados em outras áreas, como por exemplo na Saúde, bem antes desse período.

O trabalho de Kitchenham (2004) é considerado um marco no uso de estudos secundários na área de Engenharia de Software. No trabalho, a autora propõe precedimentos sobre como executar uma RS e aponta os benefícios que a técnica pode trazer para trabalhos científicos. Biolchini *et al.* (2005) traçam um paralelo entre as revisões bibliográficas realizadas nas áreas da Medicina e da Engenharia de Software. Os autores discutem o processo de RS e propõem etapas e *templates* para a execução dessas etapas dentro de RS em Engenharia de Software, além de proporem que o empacotamento dos dados deve ser realizado em todas as suas etapas.

Brereton *et al.* (2006) apresentam lições aprendidas ao aplicar RS em Engenharia de Software que vão desde a qualidade dos estudos encontrados até a síntese da pesquisa. Afirmam que, dado o crescente número de resultados experimentais, faz-se necessário adotar técnicas sistemáticas para obter resultados balanceados e objetivos. Também criticam a qualidade dos resumos e títulos dos estudos primários na área de Tecnologia de Informação (TI), a qual inclui

Engenharia de Software. Segundo os autores, os textos são ruins e com falta de informação necessária para um melhor entendimento do leitor quando comparados com outras áreas, o que dificulta a análise do pesquisador na busca por evidências em sua pesquisa.

Kitchenham e Charters (2007) propõem diretrizes para a execução de RS em Engenharia de Software. Os autores fazem um levantamento de como a comunidade está utilizando a técnica até então e sugerem diretrizes que deveriam ser seguidas para aumentar a qualidade da pesquisa. Staples e Niazi (2007) reportam suas experiências ao aplicar as diretrizes propostas, recomendando o seu uso, mas ressaltando algumas dificuldades encontradas. Como resultado, expõem a necessidade da questão de pesquisa ser elaborada de maneira clara, apoiam a reexecução de RS para reforçar os resultados e levantam a necessidade de criar um repositório de RS para a área de Engenharia de Software, como já existem em outras áreas. Dybå e Sjöberg (2008) realizam uma pesquisa com relação ao impacto das evidências dentro da RS em Engenharia de Software. Segundo os autores, uma RS é tão boa quanto as evidências que compõem a mesma.

MacDonell et al. (2010) conduzem em seu trabalho uma avaliação de quão confiáveis são as RS conduzidas em Engenharia de Software. Os autores concluem que pesquisadores com a mesma questão de pesquisa, mesmo sendo de diferentes grupos de pesquisa e realizando a revisão com diferentes métodos, chegam a resultados similares. Desta forma afirmam que o método de RS é robusto e confiável, desde que executado corretamente. Zhang *et al.* (2010) discutem estratégias para a busca de estudos bibliográficos em Engenharia de Software durante a execução de estudos secundários. Fabbri *et al.* (2013) discutem a iteratividade da condução de uma RS, externalizando pontos de iteração de suas fases com base na experiência e relatos de pesquisadores da área.

Em relação ao outro tipo de estudo secundário, o mapeamento sistemático, Petersen *et al.* (2008) propõem diretrizes para a execução desse novo tipo de estudos secundários em Engenharia de Software. O novo tipo foi proposto para as pesquisas que avaliavam os estudos primários mais superficialmente. O método proposto foi bem aceito pela comunidade e aplicado em diversas pesquisas nos anos subsequentes. Uma atualização das diretrizes para a condução de MS na Engenharia de Software é proposta por Petersen *et al.* (2015). No trabalho, os

autores atualizam as diretrizes anteriormente propostas e sugerem a inclusão da técnica *snowballing* em MS.

Jaleli e Wohlin (2012) fazem uma comparação entre a execução de estudos secundários utilizando a técnica de busca tradicional e outra utilizando uma técnica baseada em referências (*snowballing*). Os autores concluem que ambas as técnicas encontram resultados relevantes. Dentre os resultados, alguns são encontrados em comum e outros de forma complementar.

Marshall *et al.* (2014) avaliam ferramentas para a execução de RS. Nessa mesma direção podem ser encontrados outros trabalhos como por exemplo (Al-Zubidy e Carver; 2014) e (Marshall e Brereton; 2015), o que demonstra o interesse da comunidade por obter apoio computacional na execução de estudos secundários, para tentar amenizar a complexidade envolvida em sua execução, sobretudo na busca e avaliação de estudos primários.

Não é simples identificar os estudos primários que devem, ou não, compor uma pesquisa bibliográfica. Para identificá-los, existem diversas técnicas de busca por evidências bibliográficas que são executadas durante uma ou mais etapas ao conduzir uma pesquisa. Em um cenário considerado ideal, um pesquisador deve escolher a técnica de busca de estudos que, em tese, permita identificar todos os estudos relevantes sobre um determinado tema de pesquisa. Na prática, é muito difícil identificar e avaliar todos os estudos existentes tendo em vista a enorme quantidade disponível e a dificuldade para encontrá-los. A técnica de busca escolhida deverá ser aquela que permita identificar a maior quantidade de estudos com o menor esforço possível (Dieste e Padua, 2007), mas isso nem sempre é uma tarefa simples para o pesquisador. Considerando essa dificuldade, existem técnicas comumente empregadas nos estudos secundários, dentre elas: busca com *strings* (*string search*), bola de neve (*snowballing*), busca por literatura cinzenta (*gray literature search*), opinião de especialistas (*expert opinion*), autores relevantes e grupos de pesquisa (*research group search*) e busca manual (*hand search*).

Algumas dessas técnicas não são formalizadas na literatura, mesmo sendo amplamente empregadas para levantamento bibliográfico. Muitas vezes, o pesquisador faz uso de várias delas para atingir os resultados desejados. No contexto deste trabalho, a técnica é nomeada de acordo com a sua característica principal ou pelos estudos alvos do método de busca, de modo a agrupá-las para um melhor entendimento. É importante ressaltar que, muitas vezes, não existe uma

divisão clara entre as técnicas e, em uma pesquisa, nem sempre é possível separá-las. As técnicas são aplicadas diversas vezes em um estudo secundário de forma exaustiva pelos pesquisadores que desejam obter melhores resultados.

Dentre as técnicas mencionadas, *snowballing* é uma das principais. Essa técnica, embora seja menos utilizada que a busca com *strings*, é empregada para melhorar os resultados encontrados em estudos secundários. Sua utilização ocorre predominantemente de forma intuitiva e existem poucos trabalhos sobre como conduzir estudos secundários utilizando *snowballing*. Recentemente a técnica tem conquistado um espaço maior e sendo mais discutida pela comunidade. Wohlin (2014) propõe diretrizes para a execução da técnica *snowballing*. Ainda assim, é importante destacar que existem diversos desafios e lacunas que devem ser trabalhados para potencializar ainda mais a técnica.

Um dos principais desafios é como definir critérios de parada quando a técnica *snowballing* é aplicada. Critérios de parada podem ser entendidos como critérios de encerramento da coleta de estudos primários. A tarefa de definir quando o conjunto de estudos encontrado é suficiente para a pesquisa pode ser complexa e depende do tema de pesquisa escolhido. Wohlin (2014) discute possibilidades formalizando diretrizes para a execução da técnica, entretanto existem lacunas que não são preenchidas em seu trabalho. Não existe consenso da comunidade com relação à sua execução e a quando encerrar a busca por novos estudos, sendo o segundo ponto o tema central de discussão neste trabalho.

Badampudi *et al.* (2015) e Wohlin (2016) apresentam considerações sobre a pesquisas com a utilização de *snowballing* em estudos secundários. A técnica mostra uma eficácia superior à busca com *strings* em diversos casos estudados pelos autores. Nos trabalhos de Felizardo *et al.* (2016) e Wohlin (2016), o uso de *snowballing* como mecanismo para atualizar estudos secundários é avaliado, sendo esse tipo de busca também efetivo para atualizar revisões previamente construídas com *strings* de busca.

Em relação ao suporte computacional para execução de *snowballing*, existem algumas ferramentas que auxiliam nesse sentido. Porém, os recursos existentes ainda são limitados e não abrangem todo o potencial da técnica, limitando assim as ações do pesquisador que a executa. Nenhuma das ferramentas é utilizada em larga escala, sendo algumas apenas propostas em trabalhos acadêmicos. Para nenhuma

das ferramentas encontradas um estudo elaborado foi feito e, em todas elas, os critérios de parada são definidos manualmente pelo pesquisador.

1.2 Motivação e Objetivos

Dado o contexto apresentado, no qual é destacada a importância da identificação do maior número possível de estudos primários relevantes para os estudos secundários, e como pode ser difícil e trabalhoso encontrá-los, a técnica *snowballing* tem-se mostrado importante nesse sentido. E, como há uma grande demanda de esforço para localizar estudos primários por meio de *snowballing*, o trabalho aqui proposto visa a contribuir com o pesquisador que opte por aplicar a técnica *snowballing*, procurando reduzir o esforço empregado, mas mantendo a qualidade dos resultados obtidos.

Assim, o objetivo deste trabalho é propor critérios de priorização de estudos primários identificados por *snowballing* tendo por conjunto inicial de estudos os resultados obtidos com a aplicação de buscas com *strings* em máquinas de busca, de modo a tornar a execução da técnica mais eficiente do que a execução do *snowballing* completo (que verifica todas as referências), mas mantendo a qualidade dos resultados. Além disso, para tentar minimizar o esforço requerido para aplicar a técnica, o apoio computacional torna-se bastante útil. Assim sendo, como objetivo secundário, temos a implementação de recursos para executar a técnica *snowballing* e aplicar os critérios de priorização propostos na ferramenta StArt (Fabbri et al. 2016), uma ferramenta de apoio à condução de estudos secundários, criada e mantida no Laboratório de Pesquisa em Engenharia de Software (LaPES) da Universidade Federal de São Carlos (UFSCar).

1.3 Metodologia de pesquisa

Para realizar o trabalho proposto, a pesquisa foi dividida em três etapas. Na primeira etapa foram feitos estudos bibliográficos visando a identificar o estado da

arte relacionado a *snowballing* dentro da Engenharia de Software e em algumas outras áreas de pesquisa tais como Medicina e Saúde. Lacunas existentes no que diz respeito à execução de *snowballing* foram detectadas nessa etapa e serviram de base para a formulação de possíveis melhorias na aplicação da técnica.

Na segunda etapa, foram elencadas possíveis estratégias de execução de *snowballing*. Depois foram formulados os critérios de priorização de referências e estratégias para redução do esforço do pesquisador que aplica *snowballing*. Então foi conduzido um estudo para identificar dificuldades na aplicação de *snowballing* e foi desenvolvido um protótipo de *software* que viabilizou a execução de estudos sobre as estratégias propostas.

Na terceira etapa, foram conduzidos estudos para a avaliação do trabalho. Primeiramente, foi feito um estudo exploratório para avaliar o suporte computacional desenvolvido, sobretudo no que diz respeito aos algoritmos para extração e identificação de referências com base em arquivos BibTex e em textos completos dos estudos. Por fim, foi realizado um estudo experimental considerando a aplicação dos critérios de priorização de referências propostos neste trabalho.

1.4 Organização do Trabalho

Este trabalho está organizado em seis capítulos. No Capítulo 1 foi apresentado o contexto no qual este trabalho está inserido, bem como os objetivos principais.

No Capítulo 2 é apresentada a fundamentação teórica, descrevendo os tipos de estudos secundários utilizados na Engenharia de Software e as técnicas de busca empregadas para recuperação de estudos primários.

No Capítulo 3 são descritos os estudos secundários conduzidos referentes à técnica *snowballing*.

No Capítulo 4 são apresentadas estratégias para a aplicação de *snowballing*, os critérios de priorização de referências propostos e o apoio computacional desenvolvido para apoiar a execução de *snowballing* em estudos secundários.

No Capítulo 5 é apresentado um conjunto de estudos experimentais realizados para avaliação deste trabalho.

No Capítulo 6 são apresentadas as conclusões, contribuições, lições aprendidas e oportunidades futuras.

Por fim, no Apêndice A é descrito o protocolo do MS apresentado no Capítulo 3, no Apêndice B é descrito o protocolo da RS também apresentada no Capítulo 3, no Apêndice C é descrita de forma breve a ferramenta StArt e no Apêndice D é apresentado o conjunto de estudos aceitos por Hernandez, Belgamo e Fabbri (2013), que foram utilizados em um dos estudos conduzidos, apresentado na seção 5.2 do Capítulo 5.

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma visão geral dos principais temas relacionados com o trabalho que está sendo proposto, que são: estudos secundários em Engenharia de Software e seus tipos, técnicas de busca para estudos secundários e as ferramentas de apoio computacional para aplicação de snowballing existentes.

2.1 Considerações Iniciais

Conforme mencionado no capítulo anterior, existem diversas técnicas de busca que são utilizadas na condução de estudos secundários visando a identificar material bibliográfico para análises. Essas técnicas podem ser combinadas. Os pesquisadores fazem as combinações de acordo com o objetivo da pesquisa visando a encontrar o maior número possível de fontes relevantes.

Este capítulo relata as diferentes estratégias de busca bibliográfica em estudos secundários, dando ênfase maior para a técnica *snowballing*. É apresentada uma visão geral sobre os tipos de estudos secundários, bem como as diferentes técnicas de busca empregadas, e também como a técnica *snowballing* é utilizada pela comunidade científica, incluindo uma discussão sobre ferramentas para estudos secundários, mais especificamente na área de Engenharia de Software.

2.2 Estudos secundários em Engenharia de Software

Como definido anteriormente, estudos secundários são conduzidos com base em um protocolo bem definido que objetivam empacotamento e replicação do processo de pesquisa. Para o propósito deste trabalho, serão exploradas as abordagens de RS e MS por serem os tipos de estudos secundários mais utilizados na área de Computação em geral e, sobretudo, em Engenharia de Software.

De forma geral, RS e MS são compostos por três fases principais: Planejamento, Execução e Sumarização. Biolchine *et al.* (2005) caracterizam as etapas e o que deve ser executado em cada uma delas pelo pesquisador.

Na fase de Planejamento o pesquisador desenvolve um protocolo de pesquisa tendo dois objetivos principais: definir as diretrizes que deverá utilizar para executar a pesquisa e documentá-la para que outros pesquisadores possam compreendê-la e reproduzi-la, se necessário.

Na fase de Execução, o pesquisador deve coletar estudos primários fazendo uso de uma ou mais técnicas de busca. Após localizados, os estudos primários deverão passar por uma ou mais etapas de seleção. Durante a etapa de seleção, o pesquisador avalia os estudos primários considerando critérios de inclusão ou exclusão definidos no protocolo da etapa anterior e deve optar por aceitá-los ou removê-los da pesquisa.

Uma RS, conforme definido em suas diretrizes, passa por duas etapas de seleção. Uma seleção inicial para remover estudos primários não relevantes, e uma segunda seleção, na qual é feita uma análise mais completa dos estudos primários, determinando o conjunto final de estudos primários que comporão os resultados da pesquisa. Algumas RS, sobre tudo em pesquisas das áreas médicas e da saúde, os estudos primários podem passar por uma pré-seleção com base apenas nos títulos para reduzir o volume encontrado. O MS, conforme definido em suas diretrizes, passa por apenas uma etapa de seleção.

Na fase de Sumarização, os dados dos estudos selecionados são extraídos e sintetizados e devem ser disponibilizados de alguma forma para a comunidade. Normalmente os resultados têm caráter mais quantitativo em um MS, podendo normalmente caracterizar uma área ou tópico de pesquisa, ou mais qualitativo em

uma RS, procurando estipular ou alterar o estado da arte de algum tópico de pesquisa.

Kitchenham *et al.* (2009) realizam uma RS para avaliar o impacto nas pesquisas em Engenharia de Software decorridos da introdução do conceito de ESBE em 2004. Segundo os autores, o número de tópicos abordados pelas RS em Engenharia de Software ainda é limitado. Dentre os estudos encontrados, a qualidade das pesquisas foi definida como satisfatória. Os autores afirmam que as evidências sugeriram que a ESBE, e consequentemente a RS, possui um bom potencial para sintetizar e tornar disponível o resultado das pesquisas. Kitchenham *et al.* (2010) conduzem uma atualização do trabalho anterior, avaliando as publicações ocorridas nesse intervalo de tempo. Segundo os autores, RS já não devem mais ser consideradas como uma técnica em estágio inicial de uso em Engenharia de Software, mas sim uma técnica amplamente difundida na comunidade, mesmo ainda não aceita como consenso. Os autores afirmam também que em alguns tópicos dentro da Engenharia de Software não é possível avaliar de forma clara a qualidade dos estudos primários envolvidos.

Seguindo a mesma premissa dos trabalhos anteriores. Da Silva *et al.* (2011) concluem que o uso de RS ainda está em estágios iniciais. Afirmam que muitas pesquisas na Engenharia de Software falham ao não avaliar adequadamente a qualidade dos estudos primários incluídos, e falham ao não descrever o método utilizado para que outros pesquisadores possam reproduzir a pesquisa. Os autores concluem afirmando que, devido a esses problemas, o potencial da técnica é reduzido.

Zhang e Babar (2012) realizam uma investigação experimental em um conjunto de entrevistas com pesquisadores que executaram RS em Engenharia de Software, na tentativa de investigar como as RS foram conduzidas. Os pesquisadores ressaltam que a técnica deve ser encarada como relativamente nova quando aplicada ao contexto da Engenharia de Software. Os autores concluem que, apesar do esforço exigido, o número de pesquisadores que optam pela técnica está aumentando. O aumento deve-se ao fato dos bons resultados que são obtidos ao aplicar a técnica.

Kitchenham e Brereton (2013) avaliam um conjunto de RS da área de Engenharia de Software com o objetivo de identificar e avaliar as pesquisas realizadas. Os autores concluem o trabalho ressaltando os benefícios do uso de

ferramentas de análise de texto nas etapas de execução das RS e afirmam que a avaliação da qualidade dos estudos primários continua sendo um dos maiores desafios durante a execução das pesquisas.

Assim, é possível dizer que as RS estão sendo cada vez mais pesquisadas e utilizadas na Engenharia de Software. A avaliação da qualidade dos estudos que compõe a revisão é um dos principais desafios ao utilizar a técnica. Ferramentas podem e estão sendo utilizadas para auxiliar a execução de RS. A técnica ainda não é aceita como consenso, embora a cada ano o número de publicações que fazem uso da técnica esteja aumentando significativamente. Isso é um indício de uma aceitação cada vez maior por parte da comunidade científica.

2.3 Técnicas de busca para estudos secundários

Um dos passos mais importantes para a execução de estudos secundários é a coleta dos estudos que comporão a pesquisa. Para identificar estudos relevantes existe uma infinidade de técnicas de busca. Segundo Deluca *et al.* (2008), criar a estratégia de busca requer planejamento, teste e refinamento. Cuidado é necessário para garantir que toda informação relevante será inclusa.

Uma das estratégias mais comuns para levantamento bibliográfico é a construção de *strings* e o uso de motores de busca que indexam bases de dados científicas. Este tipo de estratégia é a mais utilizada para identificar referências bibliográficas.

Zhang e Babar (2004) e Dickersin (1994) afirmam que, na prática, identificar estudos primários pode ser difícil por várias razões, incluindo uma estratégia de busca inadequada, heterogeneidade da linguagem que descreve os tópicos relevantes (idioma e sinônimos das palavras-chave) e o número limitado de termos indexados que descrevem a metodologia da pesquisa (limite máximo de palavras-chave permitidas nas publicações).

Segundo Choong *et al.* (2014), não é possível utilizar apenas estratégias de busca baseadas em *strings* por conta das limitações das palavras-chave pois, muitas vezes, nem todos os termos chave do tópico de pesquisa são conhecidos. Os autores afirmam que um segundo método de busca pode melhorar a pesquisa

permitindo encontrar mais resultados relevantes e não manter a pesquisa restrita a termos específicos de busca.

Deluca *et al.* (2008) propõem o uso de outros métodos como a abordagem manual de busca em periódicos, anais de conferências, avaliar a lista de referências e entrar em contato com especialistas da área. Reforçam que existem diferentes estratégias e todas podem retornar resultados relevantes para a busca, sendo assim, consideradas estratégias complementares.

As principais técnicas aplicadas usualmente empregadas em estudos secundários foram levantadas como parte desta pesquisa e são descritas a seguir.

- **Busca com *Strings* (*String Search*)**

Define-se busca com *strings* como a busca realizada em repositórios de estudos bibliográficos que indexam trabalhos de conferências, periódicos e revistas especializadas em publicações científicas. As buscas são realizadas em repositórios científicos como, por exemplo, IEEE Xplore (2015), ACM DL (2015), SCOPUS (2015), entre outros. Os motores de busca são serviços que indexam os estudos utilizando seus metadados e permitem consultas sobre o acervo indexado.

Este tipo de busca também é descrito como busca por termos chave. É formalizada como estratégia de busca padrão em RS por Kitchenham (2004) e adotada como referencial, estando presente em praticamente todos os estudos secundários na área da Computação na última década.

O principal desafio desse tipo de estratégia é garantir a adequação da *string* utilizada nas bases de dados. Define-se como *string* de busca, uma *string* composta de conectivos lógicos e termos chave. Existem diversos trabalhos científicos voltados para como construir e como refinar adequadamente as *strings* de busca para que retornem melhores resultados. Um exemplo desse tipo de estudo é o trabalho de Dieste e Pádua (2007), no qual os autores comparam diferentes composições de termos em *strings* e os resultados retornados quando aplicadas nos motores de busca.

Essa estratégia, apesar de ser a mais utilizada, possui inúmeros problemas, dentre os quais destaca-se a sensibilidade a termos chave, pois, se não forem previamente conhecidos e incluídos na *string*, podem acarretar uma pesquisa incoerente e incompleta.

Além disso, por existirem diversas diretrizes para melhorar os resultados de buscas com *strings*, é necessário tempo e esforço para que pesquisadores iniciantes no assunto possam executar a busca de forma adequada.

- **Bola de Neve (Snowballing)**

A técnica Bola de Neve (*Snowballing*) consiste em avaliar a lista de referências e a lista de citações de estudos já conhecidos para identificar novos estudos relevantes para a pesquisa. Este tipo de técnica é simples e intuitiva.

Wohlin (2014) apresenta diretrizes para a execução de buscas utilizando a técnica *Snowballing*. O autor discute e formaliza a técnica e sua execução adequada. Define ainda os termos *snowballing backward*, que consiste na coleta e análise das referências citadas pelos estudos primários, e *snowballing forward*, que consiste na coleta e análise de estudos que citam os estudos primários.

Existem vários desafios nesta técnica, dentre os quais, obter as informações complementares sobre as referências e as citações, uma vez que há pouca informação sobre uma citação na seção de referências de um estudo. Normalmente, o pesquisador tem acesso apenas ao título e informações básicas inicialmente.

Snowballing pode ser avaliada como múltiplas pesquisas sobre múltiplos autores e é frequentemente recomendada para a condução de RS (Keen et al., 2014). Isso se deve ao fato de essa técnica não ser sensível a palavras-chave e sinônimos desses termos.

- **Busca por Literatura Cinzenta (*Gray Literature Search*)**

A busca por literatura cinzenta é uma técnica que consiste na consulta a teses de doutorado, dissertações de mestrado, sites institucionais de órgãos públicos e privados com acervos de dados, livros ou qualquer outro tipo de documentação que comprove a origem e veracidade da informação apresentada. Esse tipo de fonte pode conter valor científico, porém, muitas vezes, sem o crivo de uma avaliação formal de revisores que um estudo científico publicado nos meios tradicionais normalmente recebe.

Esse tipo de material normalmente é obtido de forma manual em sites de intuições ou publicações impressas encontradas em bibliotecas por exemplo. Um

dos principais desafios é, além de identificar as fontes que disponibilizam esses estudos, garantir a confiabilidade e a veracidade das informações encontradas.

Esse tipo de pesquisa não é comum na Computação. Entretanto, em outras áreas, existem pesquisas construídas apenas com base nesse tipo de material.

Garousi, Felderer, Mäntylä (2016) apresentam o conceito de *Multivocal Literature Review* (MLR) que consiste em incluir literatura cinzenta em uma RS. Os autores afirmam que a técnica é benéfica para a execução de pesquisas e seu uso ainda é pouco difundido na Engenharia de Software.

- **Opinião de Especialistas (*Expert Opinion*)**

Essa técnica consiste em determinar quem são os especialistas no tema de pesquisa do estudo secundário e entrar em contato com eles (ou alguns deles) para pedir sugestões de referências relevantes. Isso tende a expandir o conjunto inicial de estudos que o pesquisador já conhece.

Espera-se que um especialista da área possa indicar fontes relevantes com facilidade. Contudo, neste tipo de técnica, um desafio grande é mapear quem são os especialistas para entrar em contato. Além disso, existem alguns problemas agregados, como por exemplo, o tempo de resposta e a qualidade da lista recebida dos especialistas, pois, por mais que sejam estudos relevantes, a lista sempre pode estar incompleta.

Sem um bom conhecimento prévio por parte do pesquisador que irá realizar a pesquisa, é difícil julgar a qualidade do material indicado. Na computação, essa também não é uma técnica frequentemente utilizada, mas na medicina, por exemplo, é uma estratégia de busca muito adotada.

Petersen, Vakkalanka e Kuzniarz (2015) propõe uma atualização para executar MS e recomenda consultar especialistas no tema de pesquisa que o pesquisador possa receber indicações de fontes e estudos relevantes.

- **Mapeamento de Grupos de Pesquisa (*Research Group Mapping*)**

A técnica de mapeamento de grupos de pesquisa consiste em definir os principais autores (e seus grupos de pesquisa) da área do estudo e, uma vez mapeados, cabe ao pesquisador procurar por todas as referências desses autores e

de seus colegas. Parte da seguinte premissa: se um autor publicou um estudo relevante, provavelmente terá publicado outros estudos relevantes anteriormente ou irá publica-los em algum momento do futuro.

Pode fazer uso de uma rede de colaboração, na qual são mapeadas as relações de coautoria entre os autores com publicações em comum. Essa rede pode ser usada para identificar novos autores relevantes, aumentando o número de publicações encontradas.

A técnica pode desprezar involuntariamente os autores que possuem uma única, ou poucas publicações relevantes para o tema de pesquisa em questão. Outro problema é criar viés na pesquisa considerando certos autores por terem um grande número de publicações na área, porém sem avaliar a qualidade e relevância de seus trabalhos publicados. Avaliar a qualidade dos estudos encontrados ficará a cargo do pesquisador que executa esse tipo de busca.

O desafio desta técnica é mapear corretamente os principais grupos e redes de colaboração. Esse mapeamento pode ser construído com base nos resultados das técnicas *snowballing*, busca com *strings* ou até mesmo busca manual (explicada na sequência). Esse tipo de busca está diretamente associado à técnica de *snowballing*, pois, com resultado da execução do *snowballing* obtém-se o mapeamento das redes de colaboração e grupos de pesquisa.

Petersen, Vakkalanka e Kuzniarz (2015) recomendam o uso dessa para a execução de Mapeamentos Sistemáticos.

- **Busca Manual (*Hand Search*)**

A busca manual é provavelmente a técnica mais antiga e intuitiva utilizada. Consiste em procurar todos os estudos disponíveis por quaisquer fontes, normalmente sem adotar um processo rígido e muitas vezes pouco documentado. Para este tipo de busca, faz-se o uso de trabalhos digitais encontrados na Internet e/ou trabalhos impressos encontrados em bibliotecas ou outros acervos.

Zhang e Babar (2010) e Zhang, Babar e Tell (2011) descrevem como executar buscas manuais e afirmam que combinar a busca manual com buscas com *string* permite encontrar melhores resultados.

Essa técnica permite identificar estudos que não estão nos principais veículos científicos e que ainda assim podem ser relevantes para a pesquisa. É a técnica

mais abrangente e menos sistemática de levantamento bibliográfico. Pelo fato de fazer uso de qualquer fonte de busca disponível, tem a capacidade de identificar um maior número de artefatos. O desafio é definir em quais veículos procurar, como conduzir a busca e como documentar a execução de forma mais sistemática e procedural possível para reprodução futura, caso necessário.

Apesar de não ser passível de reprodução exata e não seguir rigor científico em sua execução, a técnica é empregada na condução de estudos secundários. Sem seguir um método rigoroso de pesquisa, tenta encontrar o maior número possível de estudos relevantes ao tema de pesquisa abordado. Um problema adicional, assim como em buscas por literatura cinzenta, é como garantir a validade das fontes encontradas para que as mesmas não comprometam a qualidade da pesquisa e a força das evidências obtidas. Deluca *et al.* (2008) recomenda utilizar buscam manual com base nos locais de publicações, tornando assim o método mais preciso.

2.4 Ferramentas de apoio para estudos secundários

Existem diversas ferramentas de apoio a estudos secundários. Dentre elas, encontra um grupo menor de ferramentas específicas para RS e MS. Marshall e Brereton (2013) realizaram um mapeamento sobre essas ferramentas existentes elencando um conjunto de ferramentas para RS e MS. Marshall e Brereton (2014) e Marshall, Brereton e Kitchenham (2014) realizam uma análise dos recursos oferecidos por algumas dessas ferramentas identificadas para diversas áreas de pesquisa, incluindo Engenharia de Software.

Marshall e Brereton (2015) apresentam um catálogo online de ferramentas para estudos secundários. Existem diversas ferramentas catalogadas, permitindo realizar consultas de acordo com as características de cada uma. Uma das ferramentas catalogadas é a ferramenta StArt (Fabbri *et al.*; 2012) e (Fabbri *et al.*; 2016). A ferramenta é utilizada neste trabalho. Os recursos para *snowballing* desenvolvidos na ferramenta ainda são recentes e foram apenas apresentados e não avaliados em publicações científicas.

As ferramenta StArt, Ramani (Souza et.al; 2014) e a ferramenta de Keen *et al.* (2014) são as únicas das ferramentas apresentaram em publicações demonstrando recursos para *Snowballing* em RS ou MS. As demais ferramentas deverão incorporar recursos para *Snowballing* futuramente, entretanto, ainda não existem publicações referentes a essas ferramentas com suporte a *Snowballing*.

Recursos referentes a ferramenta StArt e relacionados a *Snowballing* são descritos no Capítulo 3 e alguns desses recursos avaliados no Capítulo 4.

2.5 Seleção semiautomática de estudos utilizando SCAS

Um dos recursos da ferramenta StArt é a seleção semiautomática de estudos na primeira etapa de seleção. A StArt realiza a seleção semiautomática por meio da estratégia SCAS (OCTAVIANO et al.; 2015). Além do estudo inicial, um segundo estudo foi conduzido (OCTAVIANO, SILVA, FABBRI; 2016) a fim de melhor avaliar os resultados.

O algoritmo aplicado na SCAS é baseado em duas principais variáveis: o *score* e o número de citações de cada estudo interno a revisão. O *score* é calculado atribuindo-se pesos para o ocorrência de cada palavra-chave da pesquisa no título, resumo e palavras chave de cada estudo. O número de citações internas a revisão, é calculado, contando o número de citação de outros estudos na RS ou MS para o estudo avaliado. O algoritmo divide os estudos em quatro grupos (quadrantes) com base nesses dois parâmetros e recomenda a classificação automática ou manual dos estudos em cada grupo se: valor de *score* elevado e com citação interna (aceita-se automaticamente), valor de *score* elevado e sem citação interna (avaliar o estudo manualmente), valor de *score* baixo e com citação interna (avaliar o estudo manualmente) e valor de *score* baixo e sem citação interna (rejeita-se automaticamente).

Para os estudos aceitos automaticamente existe uma grande probabilidade de acerto na aceitação por parte de um pesquisador que realize a seleção manualmente. Para estudos rejeitados automaticamente existe uma grande probabilidade de rejeição. Os quadrantes do algoritmo SCAS foram utilizados nas seleções de amostras no estudo 5.3 do Capítulo 5.

2.6 Considerações Finais

Estudos secundários são utilizados em diversas áreas como estratégias de pesquisa. Na Engenharia de Software existe esforço dos pesquisadores para avaliar como os estudos secundários são conduzidos. As pesquisas mostram que existem diversos pontos para serem melhorados, entre os quais, destacam-se os métodos de busca bibliográfica empregados. Neste capítulo foram descritas estratégias de busca bibliográfica utilizadas em estudos secundários. Uma ênfase maior foi dada à técnica *Snowballing*, a qual será abordada com recorrência no contexto deste trabalho, bem como o levantamento bibliográfico realizado.

Conforme sugerem as evidências encontradas, a técnica *Snowballing* é bem aceita e utilizada pela comunidade científica para a execução de busca bibliográfica dentro da etapa de identificação de estudos primários utilizados nos estudos secundários. Existem diretrizes propostas para que o pesquisador consiga executar a técnica, mas há lacunas ainda a serem preenchidas. Os principais desafios são: definir o Conjunto Inicial de estudos para aplicação da técnica, os Critérios de Expansão e os Critérios de Parada ao aplicar a técnica. A escolha equivocada de algum deles pode comprometer diretamente a qualidade, a completude e o esforço necessário para conclusão da pesquisa.

Ainda não existem muitas alternativas de suporte computacional para *snowballing* em ferramentas de RS ou MS e pesquisas nesse sentido podem auxiliar a comunidade a obter novas ferramentas que auxiliem os trabalhos existentes.

Assim, a principal contribuição deste trabalho será a determinação de estratégias para a aplicação da técnica, utilizando critérios de priorização de referências e suporte computacional para aplicá-la.

Capítulo 3

LEVANTAMENTO BIBLIOGRÁFICO SOBRE SNOWBALLING

Este capítulo descreve a realização de dois estudos secundários, sendo um mapeamento sistemático para compreender o uso da técnica snowballing na prática pela comunidade, e uma revisão sistemática para conhecer as estratégias propostas na literatura para aplicação da técnica,

3.1 Considerações Iniciais

Com o objetivo de compreender o estado da arte da técnica *snowballing*, foram conduzidos dois estudos secundários. O primeiro foi um MS visando a compreender, em linhas gerais, como a técnica *snowballing* é utilizada por pesquisadores na condução de estudos secundários. O segundo foi uma RS com o objetivo de compreender como a técnica *snowballing* é proposta na literatura. Os resultados de ambas as pesquisas e suas conclusões estão descritos no decorrer deste capítulo.

O conteúdo das duas pesquisas foi sintetizado e apenas uma parte do método empregado e os principais resultados serão de cada uma delas será evidenciado no texto. Informações referentes ao protocolo são apresentadas no Apêndice A para o mapeamento e no Apêndice B para a revisão. As *strings* de busca utilizadas, a lista completa de estudos avaliados e aceitos e demais conteúdos relacionados a estas pesquisas estão disponíveis no endereço eletrônico: <https://goo.gl/hti6wM>, e não serão descritos em detalhes no texto.

Para comprovar que a técnica *snowballing* é aceita e utilizada pela comunidade científica, foi conduzido um MS (e atualizado até abril de 2016), a partir do qual se constatou a necessidade de realizar uma RS (e atualizada até janeiro de

2017) para complementar os resultados obtidos. O MS foi executado considerando qualquer estudo encontrado que tenha utilizado a técnica *snowballing* como técnica de busca para a pesquisa, ou *snowballing* como tema de estudo. Não foram restringidas áreas de pesquisa, entretanto as *strings* de busca foram construídas apenas em inglês, português e espanhol, e as bases de dados exploradas foram escolhidas por sua popularidade em áreas da computação e/ou áreas da medicina.

Após a realização da pesquisa, foi identificada a necessidade da realização de um segundo estudo secundário, no caso uma RS, referente às publicações encontradas no MS que trataram da técnica *snowballing* como objeto de estudo. O conjunto inicial de estudos foi composto com os estudos identificados pelo MS e, a partir desses estudos, foi aplicada a técnica *snowballing*, seguido da técnica Busca Manual para identificações de novos estudos relevantes. Os resultados são apresentados nas próximas seções.

3.2 MS sobre o uso de *snowballing* em estudos secundários

Para avaliar se este trabalho seria relevante para a comunidade científica, tanto na área da computação quanto outras áreas de conhecimento, uma pergunta deveria ser respondida: *snowballing* é utilizada pela comunidade em estudos secundários? Para responder essa questão, um MS foi conduzido. Duas questões secundárias também deveriam ser respondidas ao final da pesquisa: se *snowballing* é combinado com outras técnicas por pesquisadores e se existe suporte computacional para aplicação da técnica.

O MS foi composto por estudos primários que trataram da técnica *snowballing* e estudos secundários que utilizaram a técnica *snowballing* na pesquisa. Foi realizada busca com *strings*, a qual teve por objetivo retornar MS ou RS que utilizaram *snowballing* nas buscas ou estudos primários sobre *snowballing*. O protocolo completo do mapeamento pode ser visualizado no Tabela 35 do Apêndice A.

Os estudos foram selecionados por título e resumo na seleção inicial e a extração foi conduzida, de forma geral, apenas sobre o resumo. Apenas para alguns estudos o texto completo foi avaliado. Isso ocorreu porque a grande maioria dos

estudos encontrados continham resumos estruturados ou semiestruturados e descreviam já em seus resumos a metodologia da pesquisa utilizada por seus autores, o que era suficiente para responder as questões de pesquisa. Vale ressaltar que resumos estruturados são resumos que possuem as seções bem demarcadas e contém informações de contexto, método, objetivo, resultados e conclusões delimitadas por marcadores explícitos no texto. Resumos semiestruturados são os que estão subdivididos nas mesmas seções dos resumos estruturados, mas não possuem demarcação explícita das partes. Resumos não-estruturados são aqueles que não dividem o conteúdo nos itens do resumo estruturado e, muitas vezes, são incompletos ou mal formulados. Mais informações sobre a definição de resumos estruturados podem ser encontradas no trabalho de Guimarães (2006).

Para minimizar o descarte de informação relevante, foram definidos critérios para definir quando a leitura do texto completo deveria ser executada de forma adicional à leitura do resumo. Cinco critérios foram estabelecidos e estudos que atenderam a no mínimo um desses critérios foram lidos por completo:

- i. Tratar da técnica *snowballing*.
- ii. Propor o uso de suporte computacional para *snowballing*.
- iii. Ser um estudo da área da Computação.
- iv. Usar o termo *snowballing* e não um sinônimo.
- v. Possuir o valor do *score* (atributo gerado pela ferramenta StArt) acima da média dentre os demais estudos.

Para o critério (i), se um estudo tratar a técnica *snowballing* como objeto de estudo, considerando a leitura apenas do resumo, as informações requeridas estariam incompletas, sendo, portanto, necessária a leitura do texto completo do estudo. Para o critério (ii), mais detalhes sobre o suporte computacional seriam possivelmente encontrados no corpo do texto completo. Para o critério (iii), os resumos dos estudos da área da Computação possuíam qualidade muito inferior aos resumos encontrados em relação às áreas da Saúde, principalmente com relação à descrição da metodologia, sendo assim necessária a leitura do texto completo. Para o critério (iv), o uso do termo *snowballing* indica uma preocupação em nomear o processo como definido na literatura e, conseqüentemente, o contato do pesquisador com algum formalismo relacionado à técnica, sendo necessária a leitura do texto completo também. Para o critério (v) o *score* da ferramenta StArt foi

utilizado como um mecanismo adicional para seleção dos estudos para leitura do texto completo. Esse valor consiste na multiplicação da ocorrência de palavras por pesos numéricos de relevância os quais geram uma pontuação final para cada estudo. O cálculo é puramente estatístico, porém os estudos tendem a possuir valores de *score* acima da média considerando todos os estudos. A ferramenta StArt e algumas de suas funcionalidades são descritas no Apêndice A.

Foram encontrados 2088 estudos com a execução da técnica *strings* de Busca. Com a remoção dos resultados duplicados, 1608 estudos satisfaziam a ao menos um critério de inclusão. As bases de dados pelas quais os estudos foram obtidos são exibidas na Figura 1. Ao todo, 97 estudos foram escolhidos para avaliação do texto completo, com base nos critérios (i) a (v) descritos anteriormente. A necessidade de uma RS detalhada com os estudos primários sobre *snowballing* foi constatada.

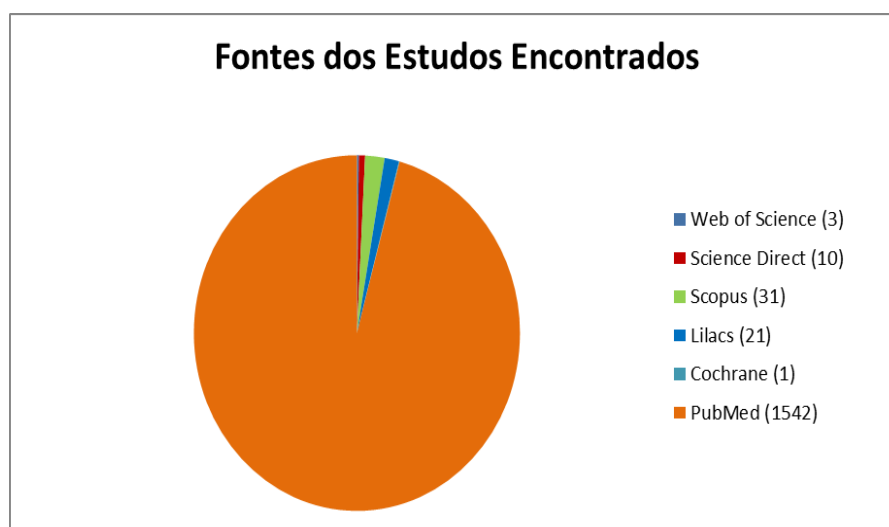


Figura 1. Base de dados em que os estudos aceitos foram encontrados

Os resultados da pesquisa são apresentados na sequência, agrupados por tópicos de interesse da pesquisa.

- **Uso da técnica *snowballing***

A técnica é utilizada pela comunidade científica, principalmente por pesquisadores das áreas da saúde. Na maioria dos casos, *snowballing* foi executada

parcialmente (apenas *snowballing backward*) e de modo intuitivo e não formalizado. A Figura 2 apresenta a relação do número de estudos que utilizaram os dois tipos de *snowballing*.

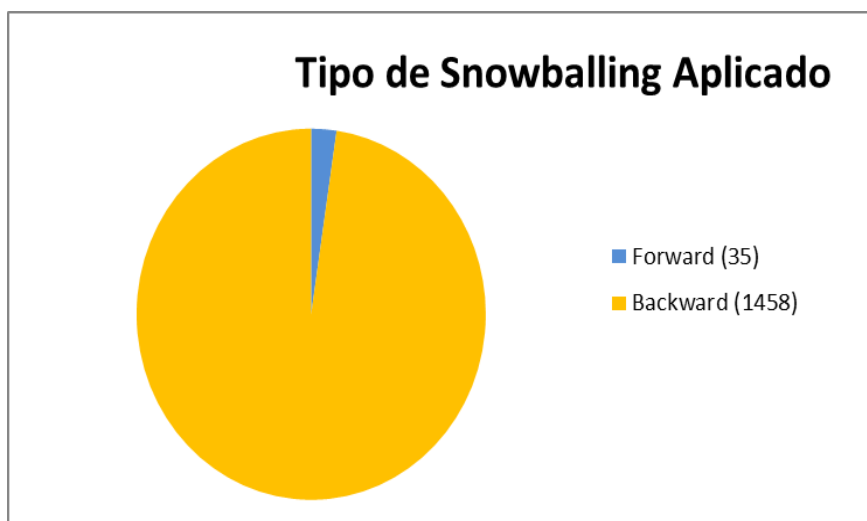


Figura 2. Número de estudos por tipo de *snowballing*

Na Tabela 1 é mostrada a relação dos estudos com o tipo de *snowballing*, a segunda coluna apresenta o número total de ocorrências para cada tipo de *snowballing*. A terceira e a quarta colunas representam respectivamente, o número de ocorrências dos estudos da área da computação e das áreas médicas e da saúde encontrados. Os dados mostram que a maioria dos trabalhos utilizam *snowballing backward* e essa afirmação, nesta pesquisa, é válida para as duas grandes áreas de pesquisa.

Tabela 1. Relação do tipo de *snowballing* aplicado por área de conhecimento

Tipo de Snowballing	Total	Computação	Áreas da Medicina e da Saúde
Snowballing Forward	35	4	31
Snowballing Backward	1458	25	1433

- **Definição do conjunto inicial**

As principais estratégias aplicadas em conjunto com *snowballing* para a construção do conjunto inicial (*seed*) foram: busca com *strings*, busca por literatura cinzenta, busca manual, opinião de especialista, autores relevantes e grupos de pesquisa. A definição e como aplicar cada uma delas é descrito no Capítulo 2.

A Tabela 2 apresenta a relação de trabalhos encontrados, suas áreas de conhecimento e os métodos de busca utilizados. Os resultados são apresentados de forma geral na segunda coluna, para a computação na terceira coluna e para as áreas médicas e da saúde na quarta coluna. A principal técnica para composição do conjunto inicial utiliza é a busca com *strings*.

Tabela 2. Relação de métodos de busca utilizados por área de conhecimento

Nome	Total	Computação	Áreas da Medicina e da Saúde
Busca com <i>Strings</i>	1590	23	1567
Busca por Literatura Cinzenta	192	3	189
Busca Manual	198	7	191
Opinião de Especialista	189	6	183
Autores relevantes e Grupos de pesquisa	5	1	4

- **Apoio computacional**

A técnica foi executada principalmente de forma manual com base na seção de referências encontrada no estudo, ou seja, foi executada a técnica *snowballing backward* manualmente.

- **Características metodológicas das pesquisas**

Os resumos dos estudos da área da saúde visivelmente possuíam melhor qualidade do que os resumos da área de computação, sendo obrigatoriamente estruturados. Os pesquisadores utilizaram diversas técnicas de busca para encontrar evidências relevantes, porém documentaram pouco a execução da estratégia de busca. Para os estudos da computação, os pesquisadores tentavam expor os procedimentos adotados de forma procedural, entretanto utilizando poucas técnicas de busca em suas pesquisas e explorando pouco os recursos menos convencionais para complementarem as suas pesquisas.

Em linhas gerais, pesquisadores da área da Saúde assumem que o processo de execução de um estudo secundário é conhecido e que pesquisadores o executaram de forma correta e adequada. Além disso, exploram diversas formas de levantamento bibliográfico. Já pesquisadores da computação normalmente utilizam

em suas pesquisas apenas uma técnica, e tentam garantir que a técnica foi executada de forma correta e abrangente documentando todo o processo e apresentado o máximo possível sobre a metodologia empregada, limitando, assim, suas pesquisas à técnica escolhida.

- **Uso de técnicas de busca em estudos secundários**

Um pesquisador da área da medicina pode buscar por testes (*trials*) de novas drogas para o tratamento do câncer em sites governamentais e de hospitais regulamentados para a execução desse tipo de testes em humanos. Esse pesquisador pode até mesmo gastar a maior parte do tempo que dispõe realizando este tipo de busca e, muitas vezes, preferindo esse tipo de pesquisa em relação a congressos e periódicos de sua área.

Provavelmente esse pesquisador entrará em contato com médicos que tenham aplicado os testes para coletar possíveis resultados e, assim, trocar experiências com seus pares. Dessa forma, o pesquisador se dispõe a usar várias técnicas para identificar literatura relevante a fim de melhorar a qualidade dos resultados.

Quando se pensa em pesquisas realizadas na área da computação e da Engenharia de Software, a maioria das pesquisas são executadas exclusivamente nas bases de dados científicas e esse comportamento se deve ao fato de que, na maioria dos casos, as principais referências na computação estão indexadas nessas fontes.

Quando se avalia a relação das técnicas de busca, observa-se que, em geral, áreas médicas predominam estudos de RS, seguidos ou não de metanálises. Como mecanismo de busca, combinam as técnicas busca com *strings* e *snowballing*, sendo comum a aplicação de outras técnicas como literatura cinzenta e opinião de especialistas.

Na área da computação, são predominantes RS e MS, com o uso de Busca por *strings* como principal técnica e, em Engenharia de Software, essa técnica é a mais empregada para a execução de RS e MS (Jalali e Wohlin; 2012). Ainda segundo os autores, em Sistemas de Informação, a técnica mais empregada é o *snowballing*.

Mais recentemente, a técnica *snowballing* vem sendo utilizada para auxiliar o levantamento bibliográfico também em Engenharia de Software, principalmente após as publicações de Jalali e Wohlin (2012) e Wohlin (2014) que, respectivamente, comprovam a complementariedade das técnicas *snowballing* e busca por *strings* e define diretrizes para a execução de *snowballing*.

Assim, não é possível afirmar que existem técnicas ótimas para a busca. Entretanto, é possível dizer que as mais empregadas são busca com *strings* e *snowballing*. As demais são incorporadas de acordo com a necessidade específica da pesquisa. Quando se observa essa comparação em relação às áreas de pesquisa, nota-se que as características das pesquisas variam em relação à área.

- **Análise temporal dos estudos**

Os estudos das áreas da Saúde foram predominantes, sendo o mais antigo do ano 2000 (*snowballing backward* apenas). Em computação o mais antigo estudo data de 2012 (já com o uso de *snowballing backward* e *forward*). A quantidade de estudos encontrados aumenta nos anos mais recentes. A distribuição dos estudos por ano pode ser observada na Figura 3.



Figura 3. Distribuição temporal do número de publicações encontradas

- **Conclusão**

Uma grande quantidade de estudos foi avaliada e a necessidade de uma pesquisa adicional foi constatada. A técnica *snowballing* é utilizada pela comunidade científica, sendo aplicada por um grande número de pesquisadores. Existem poucas fontes bibliográficas que tratam de *snowballing* como objeto de estudo.

Com base nos dados obtidos, é possível observar que a técnica de busca mais utilizada é *string* de busca, cujos resultados são ainda melhorados com a aplicação posterior de *snowballing backward*.

Existem ferramentas computacionais precursoras que estão sendo desenvolvidas para auxiliar a execução da técnica, porém, ainda são superficiais e pouco utilizadas na prática. A comunidade científica, em geral, costuma combinar a técnica *snowballing* com busca com *strings* para melhorar os resultados da pesquisa. A técnica mostra-se eficaz, mas demanda um esforço grande por parte do pesquisador e por isso ainda não é utilizada em todas as pesquisas realizadas pela comunidade científica, mesmo provando-se eficaz em diversos contextos.

3.3 RS sobre *snowballing* aplicado em estudos secundários

Uma RS foi conduzida para os estudos primários sobre *snowballing* encontrados no MS anterior (não foram executadas buscas com *string* na RS). Esta pesquisa foi necessária para complementar o MS realizado, procurando esclarecer alguns pontos que não foram detalhados com profundidade na pesquisa anterior. O protocolo completo da revisão pode ser visualizado no Tabela 36 do Apêndice B.

O objetivo dessa nova pesquisa foi caracterizar a técnica *snowballing* como é descrita na literatura, identificar potenciais lacunas a serem trabalhadas, desafios existentes, e detalhes de suporte computacional para auxiliar na aplicação da técnica. Diferentemente da pesquisa anterior, todos os estudos foram lidos na íntegra, isto é, os textos completos dos estudos foram analisados.

Foi aplicado *snowballing* no conjunto inicial de estudos. Os novos estudos encontrados foram avaliados preliminarmente com base no título. Sendo eles considerados potencialmente relevantes, então o resumo de cada estudo foi localizado para realizar a etapa de seleção inicial do estudo. O mesmo procedimento foi executado em todos os estudos aceitos até que nenhuma fonte relevante fosse mais encontrada.

Os resultados da pesquisa são apresentados na sequência, agrupados por tópicos de interesse da pesquisa.

- **Formalização da técnica em acordo com a literatura**

Segundo Keen *et al.* (2014) a eficácia do método de *snowballing* faz com que a mesma seja a melhor prática para revisões sistemáticas, porém é uma característica do método ser demorado e tedioso.

A técnica *snowballing* consiste em, para um dado estudo, identificar novos estudos a partir da lista de referências e da lista de citações. Alguns pesquisadores utilizam como sinônimo de *snowballing* o termo *Snowballing Sampling*. No entanto, a segunda nomenclatura pode causar confusão com uma técnica de amostragem estatística muito utilizada na coleta de *subjects* para estudos científicos e, portanto, recomenda-se que não seja utilizada. *Citation Track* e *Pursuing References List* são sinônimos de *snowballing* completo ou parcial e, normalmente, representam apenas *snowballing backward*.

Segundo Wohlin (2014), para a execução da técnica *snowballing* em estudos secundários, a primeira etapa é a definição do conjunto inicial, denominado conjunto *start* ou conjunto *seed*. Este conjunto é uma coletânea de estudos nos quais a técnica será executada inicialmente.

Segundo o autor, existe uma série de cuidados que devem ser tomados na escolha dos estudos do conjunto inicial, pois a composição desse conjunto terá impacto direto na qualidade, volume e abrangência dos novos estudos encontrados.

Uma vez definida a lista inicial, para cada um dos estudos contidos na lista, devem ser executadas as técnicas *snowballing backward* e *snowballing forward*. A cada vez que esse procedimento é realizado para todos os elementos da lista, e é dito que uma iteração ocorreu. E para o resultado de cada iteração pode-se aplicar novamente nos novos estudos encontrados, o que resulta em novos resultados. Após repetir o processo, o pesquisador deve definir algum critério de parada. Definir o critério de parada não é simples e pode implicar em uma série de consequências.

Nem todos os estudos da lista de uma iteração devem necessariamente ser incluídos no *snowballing*. O pesquisador tem a possibilidade de escolher os mais relevantes, ou utilizar um critério próprio, reduzindo assim o esforço empregado. Os estudos podem ser classificados de forma iterativa à medida que vão sendo identificados pela técnica *snowballing*, ou isso pode ocorrer após todos os estudos serem encontrados, decisão que cabe ao pesquisador. As demais etapas dos estudos secundários prosseguem normalmente após a execução do *snowballing*.

- **Matriz de citação e grafo de citação**

Wohlin (2014) propõe o uso de uma matriz de citação para auxiliar na aplicação de *snowballing*. Uma das dimensões da matriz representa os estudos citados e a outra os estudos citantes. A matriz cresce proporcionalmente à medida que novos estudos são identificados. Um exemplo de matriz de citação para uma pesquisa real é apresentado no trabalho de Badampudi, Wohlin e Petersen (2015). A partir da matriz é possível construir o grafo de citação (ou grafo de referências das referências) apresentado por Fabbri *et al.* (2012), que é uma representação em forma de grade da matriz de citações.

- **Suporte computacional para *snowballing***

Foram identificadas três ferramentas específicas para a execução de *snowballing* em estudos secundários e que aplicam a técnica de forma completa (*snowballing backward* e *snowballing forward*): a ferramenta proposta por Keen *et al.* (2014), a ferramenta RAMANI, apresentada por Sousa *et al.* (2013) e Bezerra *et al.* (2014) e a Ferramenta StArt apresentada em Fabbri *et al.* (2016).

Keen *et al.*, (2014) discutem requisitos para a execução automática de *Snowballing* por meio de suporte computacional e, assim, uma ferramenta é apresentada, porém não nomeada. O objetivo central do trabalho dos autores foi definir um método para identificar o resumo e texto completo de citações para os dados encontrados na seção de referência do estudo. Os autores afirmam que, ao automatizar o processo de *Snowballing*, pode ser possível reduzir o tempo e esforço das tarefas de coleta e avaliação de dados, tais como manter a pesquisa atualizada até a data corrente.

Segundo os autores, os principais desafios técnicos para a execução de *snowballing* são: obter o texto completo das citações encontradas, reconhecer as *strings* de citação dentro do texto e recuperar novas citações a partir das citações já encontradas. Tudo isso torna o processo tedioso e consome muito tempo. Com a rápida taxa de crescimento de estudos a serem avaliados, a automatização da técnica pode reduzir significativamente o tempo dedicado para a busca bibliográfica. Os autores propõem uma abordagem composta de quatro passos para a execução

de *Snowballing*: (i) extrair as *strings* de citação do texto completo; (ii) localizar as citações; (iii) dividir o texto das citações; e (iv) reproduzir o processo recursivamente nas novas citações encontradas.

Outra ferramenta proposta na literatura é a RAMANI, uma ferramenta colaborativa para a condução de MS. É apresentada por Sousa (2013) e Bezerra (2014). A ferramenta permite a aplicação de *snowballing forward* e *Backward*, sendo os estudos identificados por seus DOI (*Digital Object Identifier*), que é um identificador digital único que alguns estudo possuem. A execução do *Snowballing* é possível a partir de um *Web Crawler* que extrai as informações das páginas da (ACM DL, 2015). Um *Web Crawler* é um software que procura informações específicas em uma página *Web*, e tem a capacidade de seguir links e processar o texto HTML. Exemplo de *Web Crawler* são o GoogleBot e o BingBot, componentes de software dos buscadores Google (Google Search, 2015) e Bing (Bing, 2015), respectivamente, para indexar os resultados.

É possível mencionar que essa abordagem possui dois pontos negativos. O primeiro é que, se a ACM (ACM DL, 2015) mudar a página para a qual o *Web Crawler* foi desenvolvido, implicará no não funcionamento da ferramenta. O segundo ponto é que o uso desse tipo de aplicação pode ser considerado um ataque, por ser interpretado como ação robô (*Bot*). Ação robô é toda ação realizada por uma aplicação computacional automaticamente, sem intervenção humana, e que consome os recursos de uma aplicação, o que, muitas vezes, é proibido nos termos de uso das máquinas de busca. Dessa forma, esse tipo de aplicação normalmente não suporta um número elevado de usuários simultaneamente.

A ferramenta StArt permite a execução de *snowballing backward* por duas técnicas distintas: extração do texto completo e pelo arquivo BibTex (2016). A ferramenta também permite a importação dos resultados de *snowballing forward* de forma manual. Os recursos presentes na ferramenta foram desenvolvidos como parte deste trabalho.

- **Comparação entre *snowballing* e busca com strings**

Um estudo foi conduzido por Jalali e Wohlin (2012) e o objetivo foi executar e comparar duas diferentes técnicas para busca de fontes bibliográficas no contexto de Engenharia de Software: a técnica busca com *strings* e a técnica *snowballing* (*backward* apenas). Segundo os autores, as técnicas retornam uma grande

quantidade de estudos em comum caso executadas de forma correta, mas também trazem resultados que são complementares entre si. Dessa forma, os autores apresentam indícios de que utilizar ambas as técnicas traria mais estudos relevantes do que executar apenas uma. Entretanto, o processo se tornará mais extenso e demorado, e portanto, a viabilidade deve ser avaliada em relação ao custo-benefício de aplicar ambas.

Os autores discutem ainda alguns resultados ao comparar as duas técnicas. Na busca com *strings*, o maior tempo é gasto em refiná-las e adaptá-las para cada máquina de busca. Já no *snowballing*, o tempo é gasto identificando as referências e citações e classificando-as até encontrar os artigos relevantes.

Um cuidado que precisa ser tomado na técnica *snowballing* é que as escolhas dos estudos são normalmente feitas com base em seus títulos apenas, pelo fato dessa informação estar mais comumente disponível na lista de referências dos estudos analisados. Um estudo que não traga uma palavra-chave relevante no título pode ser perdido ao aplicar a técnica.

Badampudi, Wohlin e Petersen (2015) realizaram um estudo comparativo entre *snowballing* (*backward* e *forward*) e buscas com *string*. Os autores concluíram que *snowballing* é tão eficiente quanto buscas com *string* e pode inclusive ser mais efetivo, contudo, depende da escolha do conjunto inicial adequado, pois caso o conjunto não seja bem construído, podem ser perdidos resultados relevantes que são encontrados em buscas com *string*.

- **Desafios para a execução da técnica *snowballing***

Para a execução e automatização da coleta de estudos pela técnica *snowballing*, existem uma série de lacunas que devem ser preenchidas. Wohlin (2014) e Jalali e Wohlin (2012) evidenciam alguns dos principais desafios para a execução de *snowballing*. É possível compreender os desafios em três categorias principais: (a) Critérios para seleção do conjunto inicial de estudos; (b) Critérios de Expansão de estudos; e (c) Critérios de Parada, ou encerramento da pesquisa, descritos a seguir.

- a. Critérios para seleção do conjunto inicial (*seed*)**

O conjunto inicial (conjunto *seed*) é o conjunto de estudos fornecidos como entrada para a execução do *snowballing*. A partir dele, novas evidências serão encontradas ao aplicar *snowballing backward* e *snowballing forward*.

A composição do conjunto inicial deve ser heterogênea e abrangente. Caso a escolha da lista não atenda a esses requisitos, ocorrerão problemas que comprometerão a qualidade da pesquisa. Segundo Badampudi, Wohlin e Petersen (2015), a eficácia de *snowballing* depende fortemente da escolha do conjunto inicial adequado.

Durante a escolha do conjunto inicial, o pesquisador deve se preocupar em incluir estudos que pertençam aos diversos grupos de pesquisa e que contenham os diversos sinônimos dos termos chave para evitar ao máximo que estudos relevantes sejam desprezados. Caso a pesquisa comece com um conjunto inicial considerando apenas um termo chave (sem considerar seus sinônimos), pode ocorrer do pesquisador, ao aplicar a técnica, não conseguir encontrar estudos que utilizem sinônimos do termo chave utilizado, sobretudo se não existirem publicações que façam a ligação entre os estudos por relação de coautoria. O mesmo comportamento ocorre para grupos de pesquisa pois, caso não existam membros dos principais grupos de pesquisa, a técnica pode nunca encontrar estudos de grupos menores. Portanto, o conjunto inicial deve ser escolhido com cuidado.

A quantidade de estudos a serem utilizados no conjunto inicial terá impacto no tamanho da lista encontrada em cada iteração de *snowballing*. Não existem pesquisas suficientes para determinar a quantidade exata de estudos que deve ser utilizada. Wohlin (2014) afirma que de 10 a 15 estudos é um número interessante, pois permite a convergência da técnica em algumas iterações.

b. Critérios de expansão

Um problema para executar *snowballing* é que seu algoritmo possui crescimento exponencial proporcional ao número de citações e referências de cada novo estudo. É praticamente impossível executar o processo em todas as referências existentes pois, a cada nova referência analisada, mais tempo é gasto pelo pesquisador. É necessário priorizar os estudos potencialmente relevantes e mais promissores. Não existe regra para determinar qual será essa prioridade uma vez que o pesquisador não tem condições de saber todas as publicações existentes

sobre o tema de sua pesquisa. Wohlin (2014) propõe executar *snowballing* em todos os estudos selecionados para a etapa de extração.

c. Critérios de parada

Não é uma questão simples definir quando a pesquisa está completa. É muito pouco provável que se encontre todas as fontes relevantes, logo deve-se priorizar algumas em detrimento a outras. Um primeiro Critério de Parada seria limitar o número máximo de estudos selecionados. Essa estratégia só é interessante quando se faz uma análise exploratória e pouco abrangente na qual a completude da pesquisa não é um fator primordial.

Outro Critério de Parada seria limitar a profundidade da busca. Quando o pesquisador executa *snowballing* em um estudo relevante, os estudos encontrados tendem a sair do tema inicial e, conforme as iterações aumentam, a tendência é que, em um dado momento, não sejam mais encontradas fontes relevantes.

Alternativa presente na literatura seria exigir que a técnica *snowballing* seja aplicada em todos os estudos aceitos pelo pesquisador. Este tipo de abordagem pode demorar a convergir para um fim.

Assim, é possível perceber que a determinação dos Critérios de Parada de aplicação da técnica é algo desafiador e influencia diretamente do tempo gasto e na qualidade da pesquisa.

- ***Snowballing* para atualização de estudos secundários**

Felizardo et al. (2016) e Wohlin (2016) executaram um estudo de caso para atualizar uma revisão previamente existente e construída por meio de buscas com *string* utilizando *snowballing forward*. A técnica mostra-se efetiva para esse fim. Wohlin (2016) ressalta que pesquisadores também podem utilizar *snowballing backward* nos resultados obtidos por *snowballing forward* a fim de obter mais resultados relevantes, entretanto, poucos novos estudos deverão ser encontrados.

3.4 Considerações Finais

Conforme sugerem as evidências encontradas, a técnica *snowballing* é bem aceita e utilizada pela comunidade científica para a execução de busca bibliográfica dentro da etapa de identificação de estudos primários utilizados nos estudos secundários. Existem diretrizes propostas para que o pesquisador consiga executar a técnica, mas há lacunas ainda a serem preenchidas.

Os principais desafios são: definir o Conjunto Inicial de estudos para aplicação da técnica, os Critérios de Expansão e os Critérios de Parada ao aplicar a técnica. A escolha equivocada de algum deles pode comprometer diretamente a qualidade, a completude e o esforço necessário para conclusão da pesquisa.

Snowballing pode ser utilizado para atualização de estudos secundários e existem poucas ferramentas desenvolvidas que apoiam a execução da técnica.

Capítulo 4

ESTRATÉGIAS PARA APLICAÇÃO DE SNOWBALLING

Este capítulo apresenta estratégias para aplicação de snowballing em estudos secundários, critérios para a priorização de referências – principal contribuição deste trabalho – e técnicas de construção do suporte computacional para melhorar a eficiência ao aplicar snowballing.

A técnica *snowballing* vem sendo utilizada por várias décadas nas áreas médicas e da saúde. Na computação é um tópico que ganhou destaque recentemente com os trabalhos de Jalili e Wohlin (2012) e de Wohlin (2014). Existem diretrizes propostas para aplicação da técnica, bem como uma série de desafios e lacunas em relação a ela, conforme será apresentado nas seções a seguir.

4.1 Subsídios para a definição de estratégias para aplicação de *snowballing por meio buscas em grafos*

Construir estratégias abrangentes e que sejam úteis em diferentes pesquisas com diferentes objetivos é uma atividade complexa, sendo difícil avaliar e mensurar a sua eficiência e eficácia dada a generalidade do contexto em que são aplicadas.

Para entender de forma mais abrangente e intuitiva as estratégias de buscas em estudos secundários, nesta seção, será apresentada uma analogia do problema de busca de referências bibliográficas e buscas genéricas em um grafo. Essa

reflexão é apresentada com o intuito de facilitar a compreensão do pesquisador com relação aos problemas para execução de *snowballing*.

Os estudos bibliográficos contidos em um estudo secundário podem ser representados em fora de grafo. Uma relação de citação entre duas publicações pode ser definida da seguinte forma: se um estudo A cita um estudo B, então existe uma aresta partido do nó rotulado como A para o nó rotulado como B em um grafo que represente os resultados da busca. Uma nuvem de citações pode ser definida como um grafo construído apenas com publicações e sem nenhuma relação de citação, ou seja, um grafo sem aresta. Quando uma pesquisa começa, ela é representada por uma nuvem de citações e, à medida que as relações de citação são mapeadas pela busca, arestas são adicionadas à nuvem.

Conforme as relações de citação são adicionadas como arestas, a nuvem de citações vai se transformando em um grafo de citações. Um grafo de citações é dito completo quando todas as relações de citação contidas dentro do escopo da pesquisa estão mapeadas nele.

Em uma pesquisa convencional, o grafo de citações está normalmente em um estágio entre uma nuvem de citação, no início da pesquisa, e um grafo de citações completo, no qual todas as citações pertinentes são mapeadas.

A Figura 4 apresenta um exemplo da nuvem de citação à esquerda (região A), do grafo de citações inicial ao centro (região B), e do grafo de citações próximo do grafo completo pertinente a pesquisa à direita (região C).

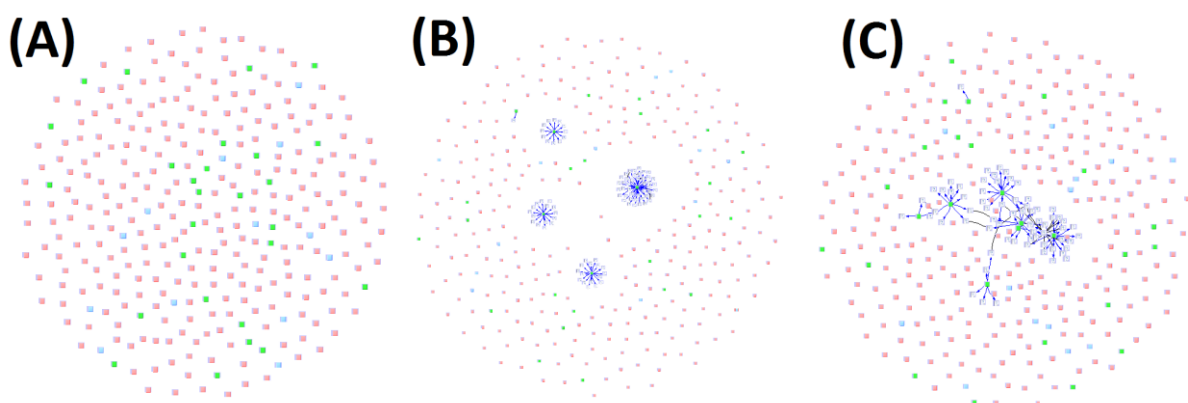


Figura 4. Exemplo de nuvem de citação (A), grafo de citação incompleto (B) e grafo de citação completo (C)

Em estudos secundários, é muito complicado obter o grafo de citações completo, dado o grande volume de estudos que podem ser necessários identificar para mapear todas as relações de citação dentro da pesquisa. Muitas vezes os pesquisadores que executam estudos secundários conhecem apenas uma parte do grafo. Se o grafo estiver completo, então, para determinar todos os estudos que citam ou são citados por outros, basta selecionar o nó equivalente no grafo e avaliar todas as arestas que se originam ou incidem nele respectivamente. Caso o grafo esteja em estado intermediário, o resultado é parcial. Essa ação equivale a consultar as referências e citações de um determinado estudo.

Agrupamentos, do inglês *clusters*, são subconjuntos pertencentes ao grafo de citações, no qual as publicações são densamente citadas entre si, e que são identificados como sub-grafos densos, ou seja, com um grande número de arestas ligando os nós a que pertencem a esse mesmo sub-grafo. A Figura 5 apresenta um exemplo com três agrupamentos distintos A, B e C em um grafo de citação.

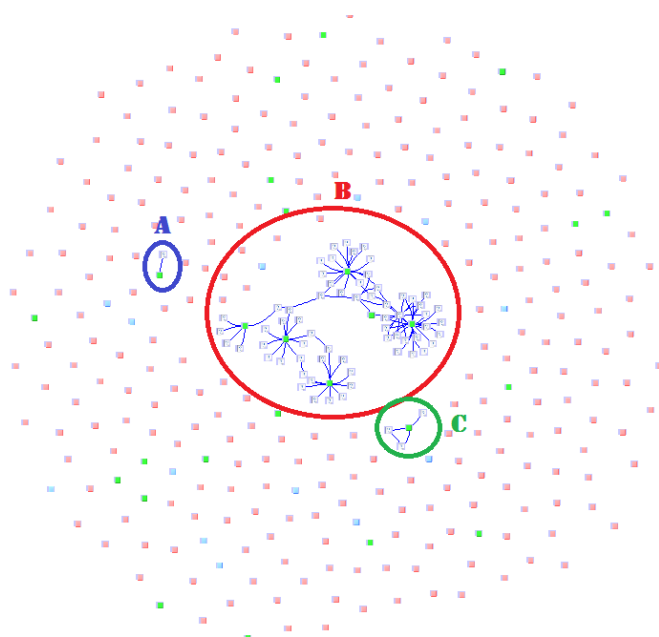


Figura 5. Exemplo de agrupamentos (*clusters*) no grafo de citação.

Com auxílio do grafo de citações e do conjunto de agrupamentos contidos nele pode-se: (i) avaliar a distribuição das fontes de pesquisa; (ii) mapear grupos de pesquisa, temas de pesquisa e linhas de pesquisa; e (iii) identificar áreas de maior incidência de trabalhos e áreas sub exploradas.

Existem dois tipos particulares de agrupamentos: por grupos de pesquisas (ou colaboração) e por sinônimos de termos chave.

Um agrupamento por grupos de pesquisa acontece quando autores dos estudos possuem um grande número de relações de coautoria com os pesquisadores dos mesmos grupos de pesquisa e suas publicações citem outros trabalhos publicados anteriormente pelos mesmos autores. Isso faz com que a parte do grafo de citações que envolve essas publicações seja densamente relacionada, representando um agrupamento (*cluster*) no grafo. Nesse cenário, no grafo de citações, é possível observar um aglomerado de estudos relacionados entre si por causa das citações e, se devidamente analisados, representam exatamente as obras de um mesmo grupo de pesquisa.

Um segundo tipo de agrupamento acontece quando existem sinônimos para um termo relacionado aos termos chave da pesquisa. Quando os sinônimos não são utilizados em conjunto pelos estudos coletados, pode não haver relações de citações que liguem grupos de estudos densamente citados, isso gera agrupamentos distintos no grafo de citações. Estudos que utilizam um desses sinônimos podem ter uma tendência a citar mais estudos que usam o mesmo sinônimo do que aqueles que utilizam outros. Um exemplo comum é o caso de duas palavras que têm mesmo significado em dois idiomas diferentes.

Os agrupamentos são importantes pois podem ocasionar uma série de problemas na interpretação do grafo de citações e na execução de métodos de busca baseados em citação e referência, principalmente para a técnica *snowballing*, uma vez que essa é uma técnica de busca que localiza estudos avaliando as arestas de citação do grafo. Dentro de um agrupamento, quando um algoritmo de busca é executado, percebe-se que muitos estudos começam a citar uns aos outros e a trazer novos estudos que também são muito citados e citam outros dentro do mesmo grupo.

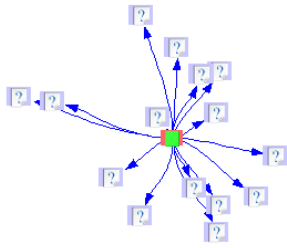
Pode-se utilizar técnicas clássicas para execução de busca em grafos a fim de buscar estudos relevantes tendo como base o grafo de citações. Ressalta-se que, nesse caso, o grafo não é completamente construído pelo fato de possuir infinitos elementos, visto que, no início da pesquisa, o conjunto de estudos existentes não é conhecido.

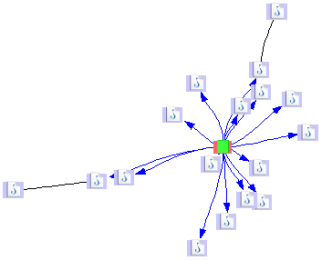
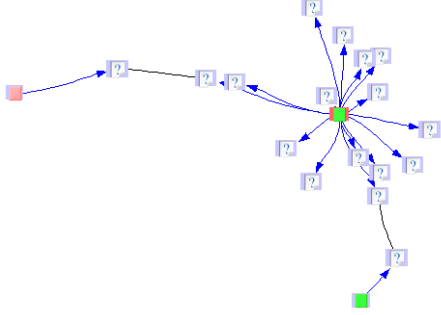
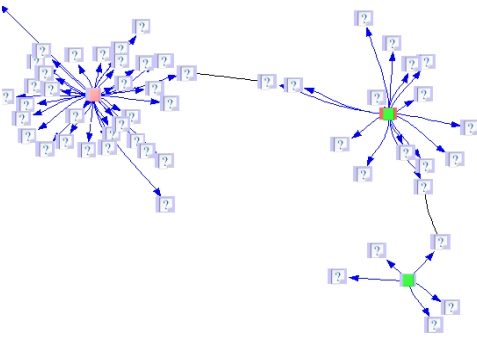
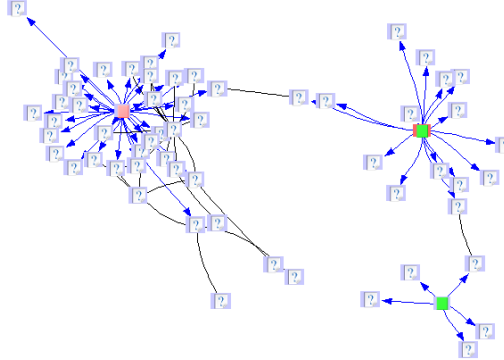
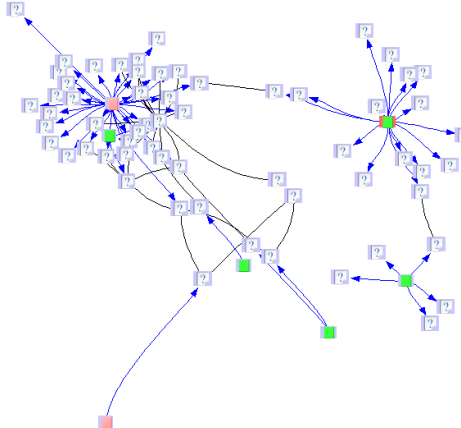
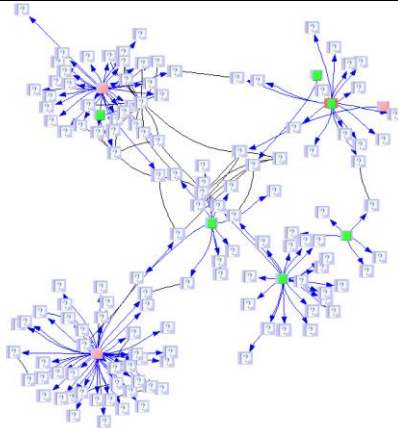
A técnica *snowballing* pode ser interpretada como busca em profundidade no grafo, ignorando o direcionamento das arestas. Caso o direcionamento das arestas

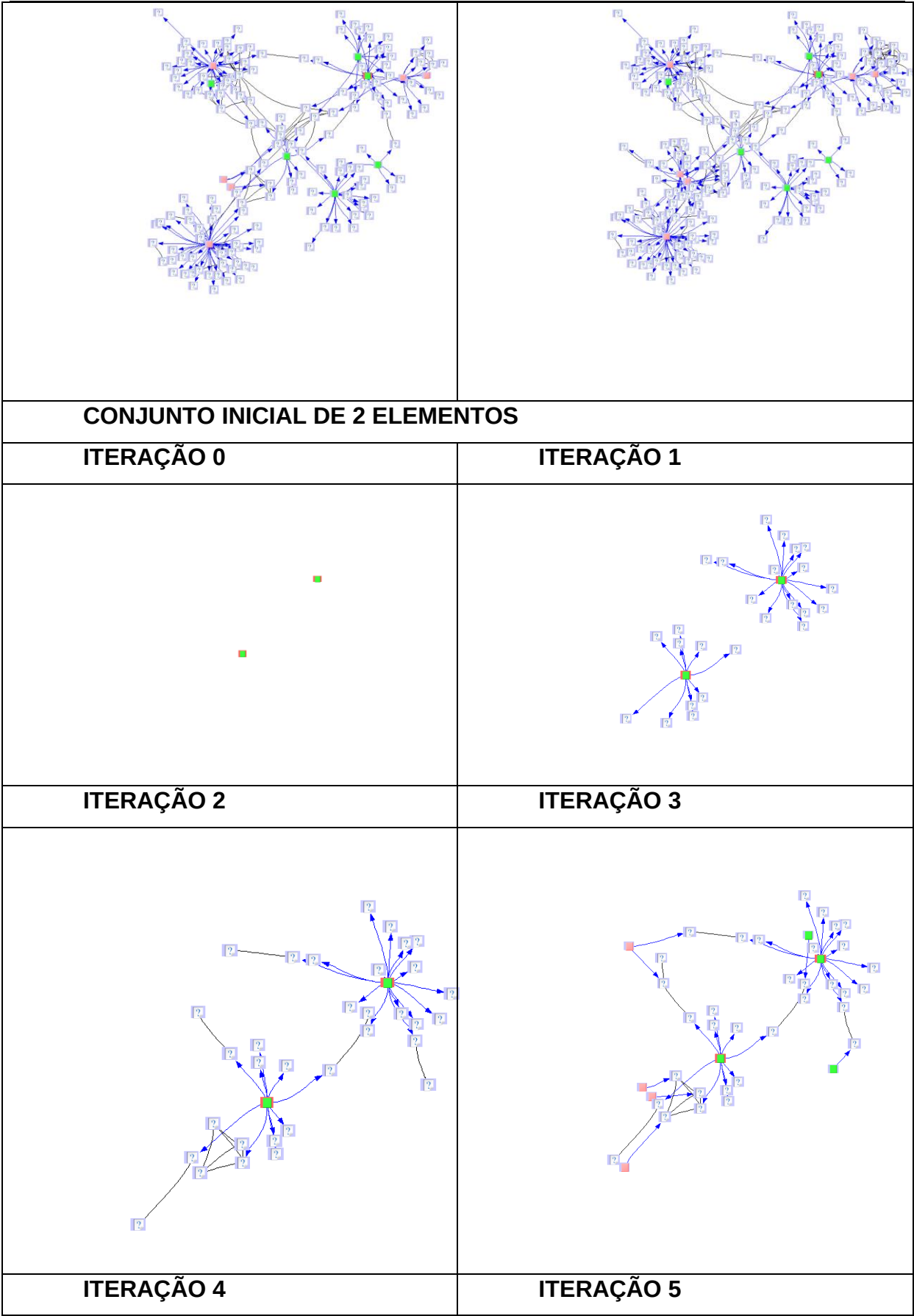
seja considerado, o método executado será parcial e representará *snowballing backward* ou *forward*, dependendo do sentido adotado.

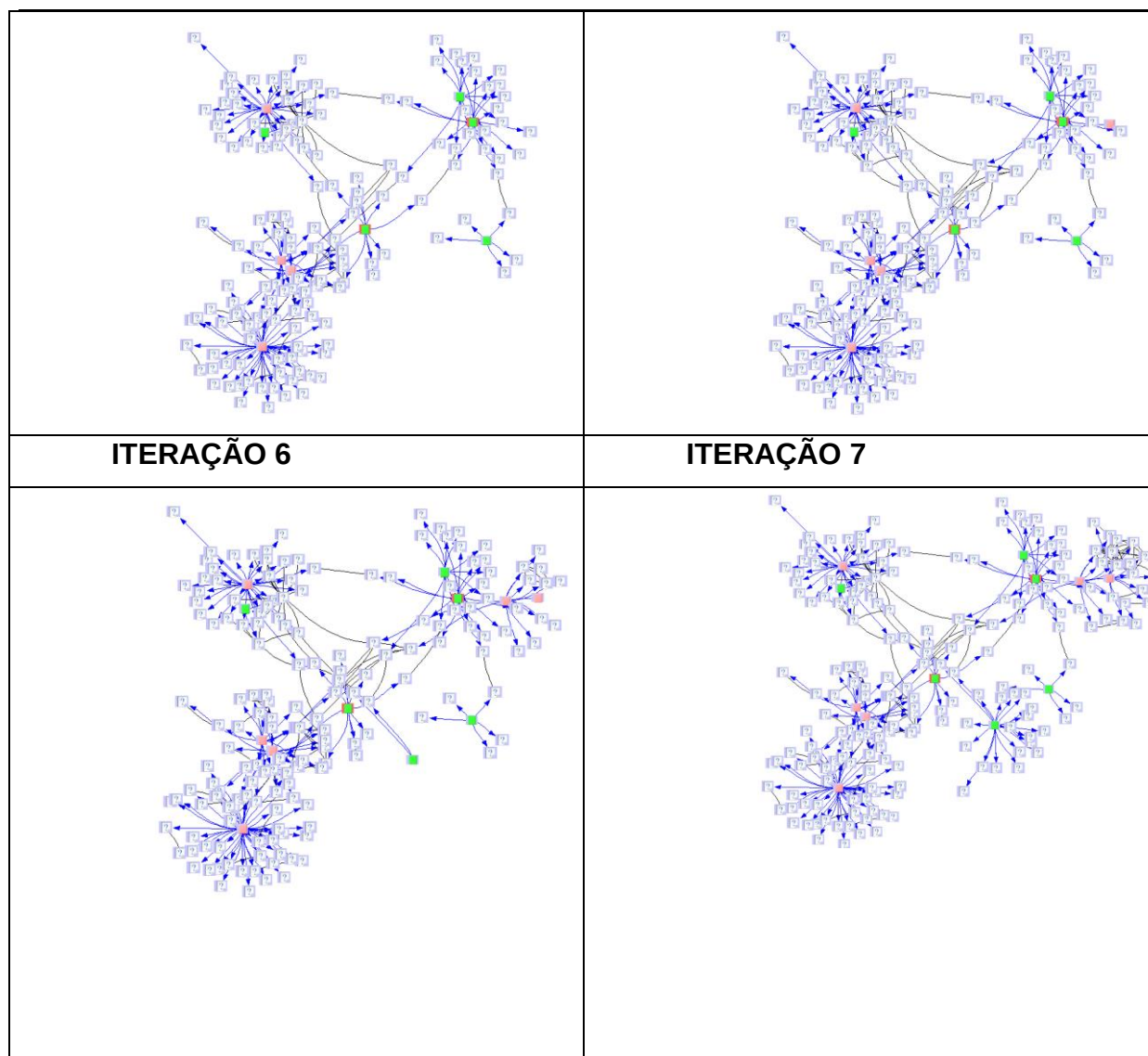
Snowballing é uma técnica que vai explorando e construindo o grafo à medida que as iterações da busca são executadas. Para que o algoritmo seja efetivo como método de busca, deve-se escolher como conjunto inicial (*seed*) da técnica *snowballing* elementos que representem os principais agrupamentos referentes ao tópico pesquisado. Entende-se por conjunto inicial o conjunto de partida para o algoritmo de *snowballing*.

No Quadro 1 é apresentado um exemplo da execução do processo de *snowballing* com dois conjuntos iniciais diferentes: o primeiro como conjunto inicial de um elemento e o segundo com dois elementos. Convergência é o estado em que, de uma interação do algoritmo para outra, o sub-grafo encontrado não se altera, ou seja, efetuando a execução da busca, nenhum novo nó (estudo relevante) é encontrado, ou seja, o grafo para de expandir. Com o exemplo apresentado no quadro, nota-se que o conjunto com dois elementos converge antes do conjunto de um elemento. A adição de um novo elemento do mesmo agrupamento ao conjunto inicial aumenta ou mantém a taxa de convergência do algoritmo, mas nunca diminui sua eficácia.

CONJUNTO INICIAL DE UM ELEMENTO	
ITERAÇÃO 0	ITERAÇÃO 1
	
ITERAÇÃO 2	ITERAÇÃO 3

	
ITERAÇÃO 4	ITERAÇÃO 5
	
ITERAÇÃO 6	ITERAÇÃO 7
	
ITERAÇÃO 8	ITERAÇÃO 9





Quadro 1 Exemplo de Iterações de Snowballing em grafos com conjunto inicial com um e dois elementos

No contexto de busca em profundidade, a convergência do método de busca significa que a busca encontrou o resultado procurado e, no caso de *snowballing*, significa que o algoritmo atendeu a ao menos um dos critérios de parada definidos pelo pesquisador ou que não haverá mais estudos relevantes para serem explorados. Não necessariamente todos os estudos relevantes foram encontrados, apenas significa que não existem mais relações de citação ou referência levando a estudos relevantes para serem exploradas. Podem existir agrupamentos que não foram identificados, por não estarem associados a elementos existentes no conjunto inicial.

O conceito de profundidade, mesmo conceito clássico utilizado em grafos e árvores, pode ser utilizado para medir o tamanho do agrupamento. Sedgewick

(2002) define bem em seu livro o conceito de profundidade como sendo a maior distância entre dois elementos quaisquer do conjunto. No caso da técnica *snowballing*, o valor da profundidade é numericamente igual ao número de iterações máximas partindo de um elemento qualquer do agrupamento para chegar a outro elemento do mesmo agrupamento, ou seja, é a distância mínima entre dois elementos.

A Figura 6 exibe o grafo de citações de uma pesquisa utilizada no exemplo do Quadro 1. Note que elementos dos demais agrupamentos ou elementos que não pertencem a nenhum outro agrupamento não foram visitados pelo método de busca. Isso ocorre pelo fato de o conjunto inicial não possuir elementos que levem os estudos pertencentes ao conjunto a esses nós. Dessa forma, publicações relevantes foram descartadas da pesquisa.

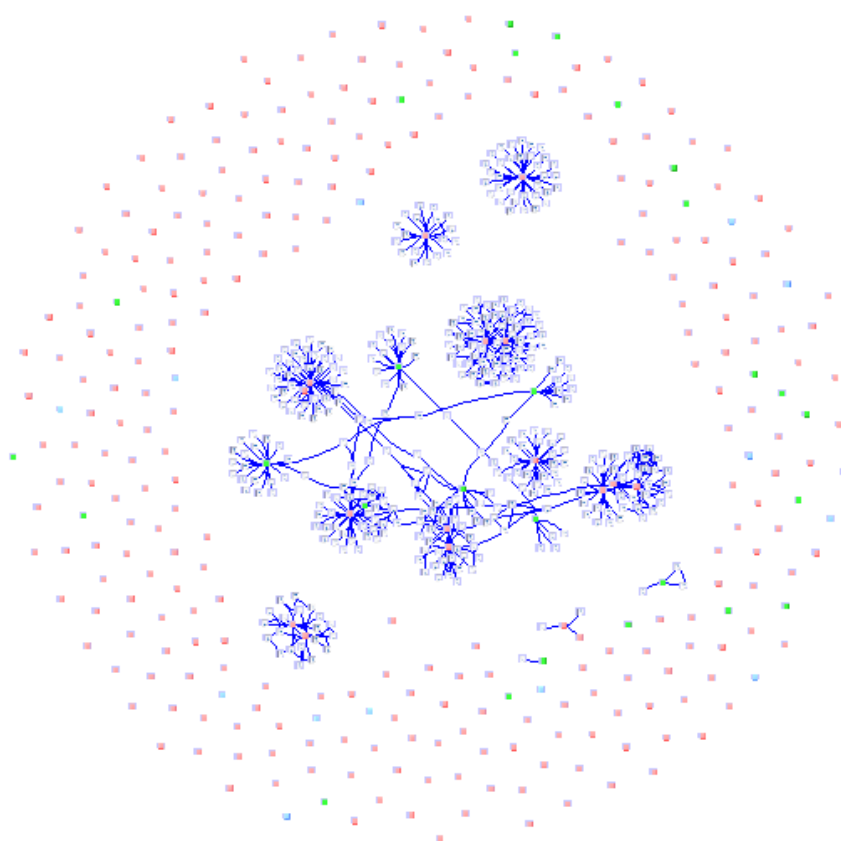


Figura 6 Grafo de citação completo gerado a partir do Quadro 1

Pelo fato de o grafo não ser finito, os dados obtidos dificilmente estarão completos. Para a sua construção, uma alternativa é o uso da matriz de citação, que é proposta na literatura para auxiliar o processo de busca de referências para estudos secundários (WOHLIN, 2014) e descrita na RS apresentada no Capítulo 3.

As imagens dessa seção foram geradas por meio da ferramenta StArt e representam dados de uma RS real, porém não publicada, sobre visualização da informação. O texto dessa seção foi descrito para auxiliar a compreensão de problemas relacionados à definição do conjunto inicial, problemas de agrupamentos e critérios de parada para *snowballing*.

4.2 Diretrizes para a execução de *snowballing* propostas por Wohlin (2014)

Wohlin (2014) descreve diretrizes para a execução de *snowballing*. Para a construção do conjunto inicial, o autor recomenda que o pesquisador construa-o com base no conhecimento prévio sobre o tópico de pesquisa.

Após os estudos processados, o autor recomenda consultar especialistas na área para receber indicações de novos estudos relevantes e realizar buscas manuais em anais de conferências e periódicos. O autor destaca que, para a construção de um conjunto inicial adequado, os seguintes pontos devem ser assegurados: (i) os estudos selecionados devem vir de diferentes grupos de pesquisa; (ii) o número de estudos que compõe o conjunto inicial não deve ser muito pequeno. Quanto mais pesquisas existirem para uma determinada área de pesquisa, maior deverá ser o conjunto inicial; e (iii) se o número de estudos já conhecidos for muito grande então o *snowballing* poderá ser executado apenas nos mais citados.

O autor propõe aplicar *snowballing* em todos os estudos aceitos pelo pesquisador e repetir o processo enquanto *snowballing* encontrar estudos relevantes. É recomendado realizar uma pré-seleção de referências considerando título, local de publicação e autores realizada manualmente. Estudos que forem considerados relevantes terão os resumos recuperados e, caso o pesquisador considere um resumo relevante, o texto completo deve ser avaliado.

De acordo com as diretrizes propostas, deve-se apenas encerrar a pesquisa quando não forem mais encontrados estudos relevantes.

4.3 Técnicas para construção do conjunto inicial

Segundo Wohlin (2014), o conjunto inicial deve ser heterogêneo. Encontrar o maior número possível de estudos para compor o conjunto inicial permite evitar os problemas de agrupamento discutidos na Seção 4.1 deste capítulo. Entretanto, o pesquisador que executa a técnica precisará de um esforço maior para compor tal conjunto inicial. As técnicas de busca apresentadas na Seção 2.3 do Capítulo 2 podem ser utilizadas para a construção do conjunto inicial.

A Figura 7 apresenta uma metáfora visual das técnicas de busca. As técnicas são complementares entre si e podem agregar resultados relevantes para a pesquisa.



Figura 7. Metáfora de técnicas de busca

Dentre as vertentes defendidas na comunidade científica de Engenharia de Software para a construção do conjunto inicial, encontram-se os trabalhos de Jaleli e Wohlin (2012), Wohlin (2014), Badampudi *et al.* (2015) que sugerem a aplicação de *snowballing* a partir de um conjunto inicial constituído pelo conhecimento prévio sobre o tema a ser pesquisado.

Existem trabalhos que propõem o uso de *snowballing* nos resultados das *strings* de busca. É o caso de Kitchenham e Brereton (2013), Webster e Watson (2002) e, ainda, trabalhos como Wohlin (2016) e Felizardo *et al.* (2016), que

propõem o uso de *snowballing forward* para atualização de estudos secundários. Nesses casos, o conjunto inicial seria composto pelos resultados de uma revisão previamente realizada.

Neste trabalho, o conjunto inicial foi fixado como busca com *string*. Não foram estudados outros tipos de variação do conjunto inicial. O conjunto inicial teve de ser fixado para permitir avaliar as estratégias e os critérios de priorização propostos.

4.3.1 Dificuldades de utilizar a busca com *strings* como *seed*

Busca por *string* é atualmente a técnica de busca mais popular e mais empregada em RS e MS na área da Engenharia de Software. Essa técnica foi proposta nos primeiros trabalhos da ESBE e presentes nos trabalhos de Kitchenham (2004) e Kitchenham *et al.* (2004), Kitchenham e Charters (2007), e Kitchenham e Brereton (2013).

Existe uma série de problemas relacionados à execução de *string* de busca. Pesquisadores iniciantes em temas de pesquisa têm dificuldades em modelar e construir as *strings* de busca para um tópico de pesquisa específico. Essa dificuldade está muitas vezes relacionada ao fato de o pesquisador iniciante não conhecer bem os termos de busca pertencentes ao domínio de sua pesquisa.

Um segundo problema em relação às buscas com *strings* é que os resultados dependem dos dados contidos nas bases de dados consultadas. Para encontrar os resultados relevantes, o pesquisador deverá mapear as principais fontes que indexam estudos para a sua pesquisa e caso não inclua alguma delas, os resultados podem ser comprometidos.

Snowballing é uma técnica que depende apenas das relações de citação e referência. Por não depender de palavras chave e nem utilizar uma base de dados de publicações específica, ao executar *snowballing* nos resultados de uma busca com *strings*, é possível encontrar resultados relevantes que não contenham as palavras chave ou que não estejam acessíveis nas bases de dados consultadas e, dessa forma, existe a possibilidade de encontrar novos estudos relevantes ainda não identificados pela *string* de busca. Por essa razão, *snowballing* pode complementar os resultados de uma busca com *strings* previamente executada.

Deve-se considerar o esforço adicional. Ambas as técnicas são trabalhosas e o pesquisador deve avaliar se vale a pena combiná-las para encontrar novos resultados relevantes.

4.3.2 Estratégias propostas para composição do conjunto inicial

A Tabela 3 apresenta técnicas propostas na literatura como forma de diretrizes (*guidelines*) ou por meio de estudos, para a determinação do conjunto inicial e aplicação de *snowballing*. A primeira coluna (#) apresenta o nome atribuído às estratégias no contexto deste trabalho. As colunas 2, 3, 4, 5 e 6 são relacionadas aos métodos de busca utilizadas nas diretrizes propostas: busca com *string* (BS), busca manual (BM), opinião de especialista (OE), buscas por literatura cinzenta (LC) e busca por autores relevantes (AR). As colunas 7 e 8 indicam se os autores sugerem o uso das técnicas escolhidas para aplicação de *snowballing backward* (SB) ou *snowballing forward* (SF). A última coluna apresenta a referência na literatura do uso de cada uma das técnicas para definição do *seed* para aplicação de *snowballing backward* ou *forward*. As referências apresentam elementos deste trabalho (I8, I9 e parte de I1) ou propostos na literatura por diversos autores que realizaram pesquisas na área (I2 a I7 e parte de I1).

Tabela 3. Técnicas para a escolha do conjunto inicial

#	BS	BM	OE	LC	AR	SB	SF	Justificativa
I1	X	X	X	X	X	X	X	MS e RS deste trabalho, Petersen et al (2015)
I2		X				X		Jaleli e Wohlin(2012)
I3		X	X			X	X	Wolin (2014), Bramaput(2015),
I4	X	X		X		X		Kitcheahm 2004
I5	X	X	X	X		X		Kitcheahm 2007
I6	X	X	X			X	X	Kitchenham and Brereton (2013)
I7	X	X					X	Felizardo et. al (2016), Wohlin (2016)
I8	X					X		Estudo da seção 5.4
I9	X					X	X	Estudo da seção 5.2

Como exemplo de interpretação da Tabela 3, podemos observar que Kitchenham (2004) (técnica I4) sugere diretrizes para utilizar busca por *string* e busca manual, e aplicar *snowballing backward* para tentar encontrar novos estudos

relevantes. Ressalta-se que a técnica 17 é proposta para a atualização de revisões previamente executadas apenas.

4.4 Critérios para aplicar *snowballing* em estudos

A técnica *snowballing* consiste em identificar estudos relevantes e, a partir deles, encontrar novos estudos potencialmente relevantes. O pesquisador deve, dado um conjunto de estudos, escolher em quais deles aplicar o algoritmo de *snowballing*.

Em princípio, uma estratégia é escolher apenas os estudos aceitos em uma revisão. No entanto, não é possível afirmar que todos os estudos que são rejeitados teriam todas as referências e citações rejeitadas. E, mesmo para os estudos aceitos, não é possível afirmar que todos possuem ao menos uma referência ou citação relevante.

Uma alternativa, visando a encontrar o máximo possível de estudos relevantes, é aplicar *snowballing* em todos os estudos identificados, sendo eles aceitos ou não. Em princípio, essa estratégia aparenta ser impraticável, entretanto, é a estratégia mais eficaz possível para aplicação de *snowballing*, embora seja pouco eficiente pois demanda um esforço excessivo para o pesquisador em troca de poucos estudos relevantes adicionais. Nesse sentido, surge uma pergunta importante e difícil de ser respondida: em quais estudos o pesquisador deveria aplicar *snowballing*?

4.4.1 Em quais estudos aplicar *snowballing*?

Wohlin (2014) sugere aplicar *snowballing* em todos os estudos aceitos no estudo secundário. Para cada novo estudo aceito, a técnica deve ser aplicada novamente. Além dessa estratégia proposta, o pesquisador pode adotar uma outra, que seria a aplicação de *snowballing* apenas nos estudos que considerar mais relevantes, reduzindo assim o conjunto de estudos inicial, ou ainda adotar outra estratégia de aplicar *snowballing* em todos os estudos, incluindo estudos rejeitados,

possivelmente encontrando mais estudos relevantes. A Tabela 4 apresenta as estratégias propostas. Na primeira coluna é apresentado o nome definido para a estratégia e na segunda coluna é descrito quando aplicar a estratégia.

Tabela 4. Relação de estratégias propostas para aplicação de snowballing

Estratégia	Em quais estudos aplicar
E1	Em todos os estudos aceitos
E2	Em todos os estudos identificados
E3	Apenas nos estudos mais relevantes

Justifica-se a proposta de aplicar *snowballing* em estudos rejeitados no seguinte cenário: um estudo pode ser rejeitado em um estudo secundário, mas servir de fundamentação teórica para diversos estudos que serão aceitos no mesmo estudo secundário. Assim, aplicar *snowballing forward* nesse estudo pode levar a outros estudos que potencialmente serão aceitos no estudo secundário. De forma similar, um trabalho futuro de um estudo que foi aceito no estudo secundário pode conter, em sua lista de referências, novos estudos que também serão aceitos no estudo secundário.

Apesar de existir a possibilidade de identificar estudos relevantes pelos estudos rejeitados, muitos estudos deverão ser avaliados para identificar poucas novas referências relevantes que talvez sejam encontradas. Técnicas computacionais podem ajudar a priorizar as referências de estudos aceitos ou rejeitados para facilitar o trabalho do pesquisador e, assim, viabilizar esse tipo de estratégia que atualmente não é aplicada pela comunidade.

4.4.2 Critérios de Priorização de Referências

Na tentativa de minimizar esforço de pesquisadores que executam *snowballing*, e principalmente dos pesquisadores que escolhem aplicar *snowballing* em todos os estudos encontrados, são propostas neste trabalho quatro técnicas para priorização de referências: (i) *score* das referências e citações; (ii) número de referências internas dentro de estudos secundários; (iii) *ranking* de conferências ou periódicos; e (iv) *ranking* de autores. Com base nesses critérios, entende-se ser possível reduzir o esforço do pesquisador sem haver perda significativa de referências relevantes para o tópico de pesquisa.

- **Score de referências (SC)**

Para ranquear os estudos por esse critério basta contar o número de ocorrências de palavras chave da pesquisa (definidas no protocolo) dentro do título dos trechos de texto referentes a uma citação ou referência encontrada. Uma vez computados os scores das referências, as mesmas podem ser classificadas em ordem decrescente por *score*.

O critério do score de referências foi inspirado no *score* da ferramenta StArt, apresentado em Fabbri et al. (2016), para ranquear estudos utilizando o resumo, título e palavras-chave, e também é utilizado por Octaviano *et al.* (2015) para realizar a seleção semiautomática de estudos primários.

- **Número de citações internas nos estudos secundários (CIT)**

Para ranquear os estudos por esse critério, basta contar o número de vezes que um estudo primário é citado por outros dentro do mesmo estudo secundário (número de citações). Uma variação desse critério é contar apenas as citações de estudos aceitos dentro do estudo secundário. Uma vez contabilizado o número de citações, basta fazer o ranqueamento dos estudos classificando-os em ordem decrescente por número de citações.

O critério do número de citações internas foi inspirado no trabalho Fabbri et al. (2012), no qual os autores afirmam quanto mais citado for um estudo por outros estudos aceitos dentro do estudo secundário, maior a chance de que ele seja relevante para a pesquisa. Octaviano et al. (2015) utiliza o número de citações internas para a seleção semiautomática de estudos primários.

- **Ranking de conferências ou periódicos (LO)**

Para ranquear os estudos por esse critério, basta contar o número de ocorrências de estudos em cada conferência ou periódico. Uma variação desse critério é contar apenas os periódicos ou conferências de estudos primários aceitos dentro do estudo secundário. Em seguida, é preciso classificar os estudos de maneira decrescente pelo número de ocorrências em conferências ou periódicos calculado para cada estudo primário.

Esse critério foi inspirado na premissa de que se muitos estudos são aceitos em uma mesma conferência ou periódico, então muito provavelmente existirão outros estudos da mesma conferência ou periódico que também serão aceitos e que ainda podem não ter sido identificados inicialmente.

- **Ranking de Autores (AU)**

Para ranquear os estudos por esse critério, basta contar o número de ocorrências de estudos primários publicados por cada autor dentro do estudo secundário. Uma variação desse critério é contar apenas os autores de estudos primários aceitos. Assim, como nos anteriores, é preciso ordenar os estudos em ordem decrescente conforme os valores calculados.

Esse critério foi inspirado na premissa de que caso um autor realize uma pesquisa em um determinado tema de pesquisa, existe uma grande possibilidade de que ele volte a publicar no mesmo tema de pesquisa. Não é via de regra, contudo autores tendem a trabalhar diversos aspectos sobre um determinado tópico de pesquisa.

É possível que, ao utilizar os critérios CIT, LO e AU, considerando os estudos aceitos, sejam encontrados melhores resultados, contudo, é necessário aplicar a seleção de uma parcela significativa de estudos antes de aplicar os critérios. Vale ressaltar que é possível que, a cada nova classificação, os valores dos critérios mudem. Essas estratégias que envolvem classificação não foram avaliadas neste trabalho.

Os critérios apresentam problemas que podem fazer com que estudos relevantes sejam menos bem ranqueados ou estudos pouco relevantes sejam mais bem ranqueados. Dentre os problemas, podemos destacar a sensibilidade a palavras chave, sensibilidade a grupos de pesquisa e sensibilidade ao ano de publicação. A Tabela 5 apresenta uma relação entre os critérios propostos e os problemas citados.

Tabela 5. Problemas relacionados aos critérios de priorização propostos

Característica	Score (SC)	Nº Referências (CIT)	Ranking de Conferência (LO)	Ranking de Autores (AU)
Sensível a Palavras Chave	SIM	NÃO	NÃO	NÃO
Sensível a Grupo de Colaboração	NÃO	SIM	SIM	SIM
Sensível ao ano de publicação	NÃO	SIM	NÃO	NÃO

4.5 Quando encerrar a pesquisa utilizando *snowballing*

Para o pesquisador, determinar quando dar-se por encerrada a busca por novos estudos relevantes é uma tarefa difícil. Ao aplicar *snowballing*, enquanto novos estudos são encontrados, recomenda-se continuar a pesquisa para identificar o maior número possível de resultados relevantes. Entretanto, dependendo do número de estudos encontrados, o pesquisador deve encerrar as buscas por novos estudos e assumir o risco de ainda existirem alguns que sejam potencialmente relevantes, mas que não foram localizados.

Não encontrar todas as referências relevantes não invalida um trabalho científico, entretanto, quanto mais relevantes forem as referências que faltam, menor será a credibilidade do trabalho perante a comunidade. Isso porque, nesse caso, é maior a chance de o pesquisador não ter levado em conta evidências importantes sobre o seu tópico de pesquisa e possivelmente desconhecer o real estado da arte sobre o tema de pesquisa. Wohlin *et al.* (2013) e Petersen, Vakkalanka e Kuzniarz (2015) afirmam que, além da quantidade de referências que faltam na pesquisa, é necessário avaliar a qualidade e relevância dessas referências para concluir se um método de busca adotado é ou não eficaz.

Para pesquisadores que utilizam busca com *string*, ao aplicar a técnica *snowballing*, a busca deverá parar de encontrar novos resultados relevantes rapidamente. Caso o pesquisador utilize *snowballing backward* para complementar a pesquisa, os resultados devem indicar estudos que não estejam presentes nas bases de dados consultadas, caso contrário pode haver problemas na construção da *string* de busca utilizada. Um exemplo é a falta de um sinônimo ou termo chave na *string* de busca. Não é recomendável finalizar a execução de *snowballing* enquanto

a aplicação da técnica ainda encontrar estudos considerados muito relevantes para o contexto do tema pesquisado.

4.5.1 Estratégias para encerrar a pesquisa realizada com *snowballing*

A estratégia mais comum para encerramento de buscas por *snowballing* é executar a técnica enquanto estudos relevantes forem encontrados e apenas encerrar a pesquisa quando não houver mais estudos relevantes, conforme as diretrizes sugeridas por Wohlin (2014).

Em alguns casos, quando os critérios de aceitação são pouco restritivos, o volume de dados encontrados pode crescer de forma exponencial. Duas alternativas para a estratégia inicial são: definir um tempo limite para a pesquisa e definir um número máximo de iterações para *snowballing*. Ambas as estratégias assumem o risco de perderem referências relevantes. Essas estratégias devem ser consideradas caso o pesquisador opte por aplicar *snowballing* em estudos rejeitados. A Tabela 6 apresenta as estratégias de parada de *snowballing* (segunda coluna) e os códigos atribuídos a elas no contexto deste trabalho (primeira coluna).

Tabela 6. Estratégias de paradas de *snowballing*

(#)	Estratégia
P1	Não encerrar a pesquisa enquanto o algoritmo de <i>snowballing</i> encontrar estudos relevantes
P2	Limitar o número máximo de iterações de <i>snowballing</i>
P3	Limitar o tempo máximo para a execução da pesquisa

4.6 Definindo uma estratégia completa para *snowballing*

Uma estratégia completa para a aplicação de *snowballing* pode ser entendida como a combinação de três elementos: um conjunto inicial, critérios de expansão de estudos e os critérios de parada da pesquisa. A Tabela 7 apresenta as opções definidas para a escolha de cada estratégia. Para definir uma estratégia completa basta combinar um termo da coluna “Conjunto Inicial” com um termo da coluna “Critério de Expansão” com um termo da coluna “Critérios de Parada”.

Tabela 7. Estratégias combinadas para *snowballing*

Conjunto Inicial	Critério de Expansão	Critérios de Parada
{I1,I2,I3,I4,I3, I4, I5, I6, I7, I8, I9}	{E1,E2,E3}	{P1,P2,P3}

Um exemplo de estratégia completa seria, por exemplo, a combinação (I7, E1, P2), que significaria: (i) adotar a estratégia de utilizar os resultados da busca por *strings* combinada com a busca manual como conjunto inicial para *snowballing* e aplicar *snowballing forward* nesse conjunto; (ii) expandir a técnica para os estudos aceitos; e (iii) limitar o número máximo de iterações que será aplicado. Outro exemplo de estratégia aplicada pela comunidade e descrita por Wohlin (2014) é a combinação (I3, E1, P1).

4.7 Apoio computacional para *snowballing*

Snowballing pode demandar um grande esforço para sua execução e o suporte computacional pode ser utilizado para auxiliar o pesquisador que aplica esse tipo de técnica. Existem desafios relacionados à construção de apoio computacional para *snowballing*, sendo os principais: coletar as referências e citações dos estudos selecionados, e classificar ou priorizar os novos estudos encontrados.

Para coletar as referências e citações por meio de um suporte computacional, é possível utilizar quatro alternativas: (i) texto completo dos estudos; (ii) arquivos de referências dos estudos; (iii) utilizar API disponibilizadas pelos motores de busca; e (iv) utilizar *web crawler*. Nenhuma das alternativas atinge todos os estudos possíveis, retornando resultados incompletos. Uma combinação das estratégias poderia trazer melhor resultado. As estratégias são descritas a seguir.

4.7.1 Extração de referências por meio do texto completo

Essa estratégia permite apenas a execução de *snowballing backward* e, ao executá-la, existem diversas dificuldades, que vão desde obter os dados a como tratá-los. Não existem algoritmos de solução exata para esse problema.

Mesmo com essas dificuldades, a técnica foi escolhida e implementada na ferramenta StArt por ser a mais flexível, uma vez que só depende do *full text* (texto

completo) do estudo, cujo *download* é um recurso quase sempre necessário para a execução da pesquisa.

Uma dificuldade dessa alternativa é criar um algoritmo independente de padrão de construção dos estudos e das normas de formatação utilizadas, por exemplo, o padrão ABNT (2016) em relação à escrita da lista de referências. O algoritmo deve reconhecer os diversos formatos.

4.7.2 Extração de referências por meio do arquivo de referências

Algumas máquinas de busca permitem exportar para um arquivo de referências BibTex (2016), além das informações dos estudos, todas as referências citadas por cada estudo. Esse recurso não está disponível em todas as máquinas de busca e, até o momento, só foi encontrado ao exportar as informações dos estudos no formato BibTex.

Os dados cadastrados nesses arquivos estão muitas vezes incompletos e sem formatação padrão, ou seja, não estão estruturados e nem utilizam normas padrão, podendo faltar dados. Essa alternativa também foi incorporada à ferramenta StArt.

4.7.3 Clientes de API Rest.

Algumas máquinas de busca possuem API (*Application Programming Interface*) de *web services*. A partir dessas APIs, é possível obter acesso a recursos que as máquinas possuem, tais como busca de estudos, download de parte do texto completo e estudos relacionados.

A dificuldade para desenvolver apoio computacional é que, para cada uma das máquinas de busca existentes, é necessário construir um cliente para a sua API, o que envolve tecnologias e arquiteturas diferentes e que sofrem constantes mudanças.

Pensando em sua utilidade prática e aplicação em cenários reais, uma grande quantidade dessas APIs deve ser desenvolvida para atender a necessidade de pesquisadores que utilizarem um *software* em seus estudos.

4.7.4 Web Crawlers

Essa alternativa consiste em implementar um *crawler* que obtenha informação a partir dos arquivos HTML da máquina de busca. O *crawler* funciona como uma ferramenta que “lê” a página web e obtém dados específicos. A leitura consiste em encontrar uma informação específica no texto HTML. O *crawler* pode ser configurado para seguir os links de um arquivo HTML e, assim, navegar entre páginas para encontrar novas informações.

Para aplicar a técnica é necessário definir quais máquinas de busca possuem as informações de referências e citações e criar *parsers* específicos para identificar as informações necessárias. O *crawler* é sensível à mudança dos arquivos HTML e, caso os mesmos mudem, pode implicar na necessidade da manutenção do software criado. Outro problema é que as máquinas de busca não autorizam execução em massa de web *crawler* e essa ação é ilegal em todas as máquinas de busca.

Neste trabalho, apenas as alternativas de extrair referências a partir do texto completo dos estudos e dos arquivos de referências dos estudos foram exploradas e implementadas na ferramenta StArt. Foram escolhidas por serem estratégias que, uma vez implementadas, demandariam poucas alterações ao longo do tempo em comparação às outras duas.

Na Tabela 8 são relacionadas algumas características (vide coluna 1) com as quatro alternativas para coleta de referências descritas anteriormente (colunas 2 a 5). As possíveis características são: configurar ação robô, o que é proibido por máquinas de busca, dependência do texto completo, dependência da máquina de busca, permitir a execução de *snowballing backward* e permitir a execução de *snowballing forward*. Para cada alternativa é mostrado se a característica descrita é relacionada a ela.

Tabela 8. Características das estratégias para a coleta de referências e citações

Característica	WebCrawler	Parser Referências	Extração de Texto	API
Configura ação robô	SIM	NÃO	NÃO	NÃO
Depende do Texto completo	NÃO	NÃO	SIM	NÃO
Depende da Máquina de Busca	SIM	SIM	NÃO	SIM
Viabiliza Snowballing Backward	SIM	SIM	SIM	SIM

Viabiliza Snowballing Forward	SIM	NÃO	NÃO	SIM
-------------------------------	-----	-----	-----	-----

Para a classificação das referências encontradas, é necessária uma revisão manual do pesquisador. Os critérios de priorização apresentados na seção 4.4.2 deste capítulo foram criados e implementados na ferramenta StArt para auxílio ao pesquisador que realiza essa tarefa.

4.7.5 Como identificar estudos em textos de referências e citações

Para viabilizar o reconhecimento dentro de um trecho de texto, foi utilizado reconhecimento por expressões regulares. Por meio das expressões, é possível identificar ano, título e autores de cada referência contida no *full text*.

As expressões possuem uma implementação interna de máquinas de estado finito e os resultados não são totalmente precisos. Ocorrem erros de reconhecimento em parte dos dados obtidos. Esses erros ocorrem quando trechos de texto, que não são de fato o trecho que deveria ser reconhecido pela expressão, condizem com os padrões implementados na máquina de estado.

Como solução adicional, para contornar os erros de reconhecimento, quando é necessário comparar estudos para determinar se tratam da mesma produção científica, são usados algoritmos de similaridade de texto.

Como alternativa ao algoritmo Salton (Salton; Allan, 1995) que é um algoritmo comumente utilizado e proposto como técnica de mineração de texto, um algoritmo foi desenvolvido pelo autor deste trabalho, o qual apresentou resultados melhores que o anterior para comparar referências. Ao considerar o erro das expressões regulares e posterior aplicação dos algoritmos de similaridade, que induzem a mais erro, esse tipo de estratégia pode apresentar resultados incorretos.

Para a implementação das máquinas de estado reconhecedoras de expressões regulares, optou-se por utilizar o reconhecimento nativo da linguagem JAVA para reconhecimento de expressões regulares (do inglês *Regex*). O mesmo utiliza uma implementação interna de reconhecimento partindo de uma linguagem de entrada. Essa tecnologia é amplamente utilizada em aplicações comerciais para validação de texto de entrada tais como e-mails, telefones e endereços.

Cada linguagem de programação possui implementações diferentes para as expressões regulares, utilizando o mesmo padrão de entrada proposto primeiramente na linguagem C.

As expressões construídas para reconhecimento de textos de referências são apresentadas na Tabela 9. A primeira coluna representa o objetivo de reconhecimento. A segunda coluna representa a expressão regular de entrada, a terceira coluna define o padrão desenvolvido para reconhecendo e, por último, a quinta coluna que representa um exemplo de *string* reconhecida.

Tabela 9. Expressões regulares para reconhecimento do texto das referências.

Objetivo	Expressão <i>Regex</i>	Padrão	Exemplo de <i>string</i> reconhecida
Ano	[1-2][0-9]{3,3}	Único	1999
Título	[,;:]*[,;:]	Padrão 1	, Introdução ao Teste de Software;
	[^"].*["]	Padrão 2	"Introdução ao Teste de Software"
Autor	(([A-Z][A-z]{1,}){1,})([,]*([A-ZÁ-Ú][]*[]*))){1,}	Padrão 1	Silva, C. ; Silva, C. R. Q.
	(([A-Z][A-z]{1,})([]*[]*[A-Z][A-z]{1,}){1,}){1,}	Padrão 2	Silva, Cleiton. ; Silva , Cleiton Rodrigo Queiroz
	((([A-ZÁ-Ú][]*[]*)[,;:]{0,1}[]*){1,}[A-ZÁ-Ú][A-z]{1,}){1,}	Padrão 3	C. Silva, C. R. Q. Silva
	((([A-Z][A-z]{1,})[]*[]*[]*){1,}[A-Z][A-z]{1,}){1,}	Padrão 4	Cleiton Silva , Cleiton Rodrigo Queiroz Silva
Referência Completa	\n+	Único	Referencia1
			Referencia2

Atualmente as expressões são utilizadas na ferramenta StArt, que seleciona uma combinação, a partir de um conjunto de combinações pré-definidas, como sendo a mais efetiva, selecionando a estratégia que identifica mais referências completas dentro do texto. Resultados completos são os dados identificados que contêm todos os principais elementos de uma citação: ano, título, autores e local de publicação.

4.7.6 Uso dos algoritmos de similaridade

A partir do uso de algoritmos de similaridade, é possível obter dois resultados importantes para a execução de *snowballing*: identificar as duplicações de estudos e identificar citações de estudos por outros estudos.

Conforme explicado anteriormente, os reconhecimentos não são exatos. Os textos reconhecidos contêm a maior parte da informação buscada, porém não

contém a informação completa. Foram testados dois algoritmos para comparação de textos próximos por similaridade: Salton (Salton; Allan, 1994) e um algoritmo desenvolvido como parte deste trabalho. O algoritmo é descrito a seguir:

Passo 1: Seleciona-se ambos os textos e remove-se todos os caracteres especiais e números deixando apenas letras de A até Z, maiúsculas ou minúsculas. Todos os espaços em branco são removidos.

Passo 2: Letras maiúsculas são substituídas pelas letras minúsculas para evitar que o algoritmo não reconheça.

Passo 3: Verifica-se se os resultados da maior cadeia de caracteres contêm os resultados da menor cadeia de caracteres.

Diferente dos outros algoritmos, que retornam um valor em porcentagem de similaridade, o algoritmo retorna verdadeiro, caso considere que os estudos tratam do mesmo trabalho, ou falso, caso contrário.

4.7.7 Interface e Visualização

Na Figura 8 é representada a interface na qual o pesquisador pode validar e interagir com os dados extraídos a partir das listas de referências obtidas pelo *full text* do estudo ou pelo arquivo exportado das máquinas de busca. Além da abordagem de extração do *full text*, foram implementados recursos para obtenção das referências pelo arquivo de citações obtido por consultas na máquina de busca. A ferramenta StArt combina os resultados gerados e os apresenta na interface.

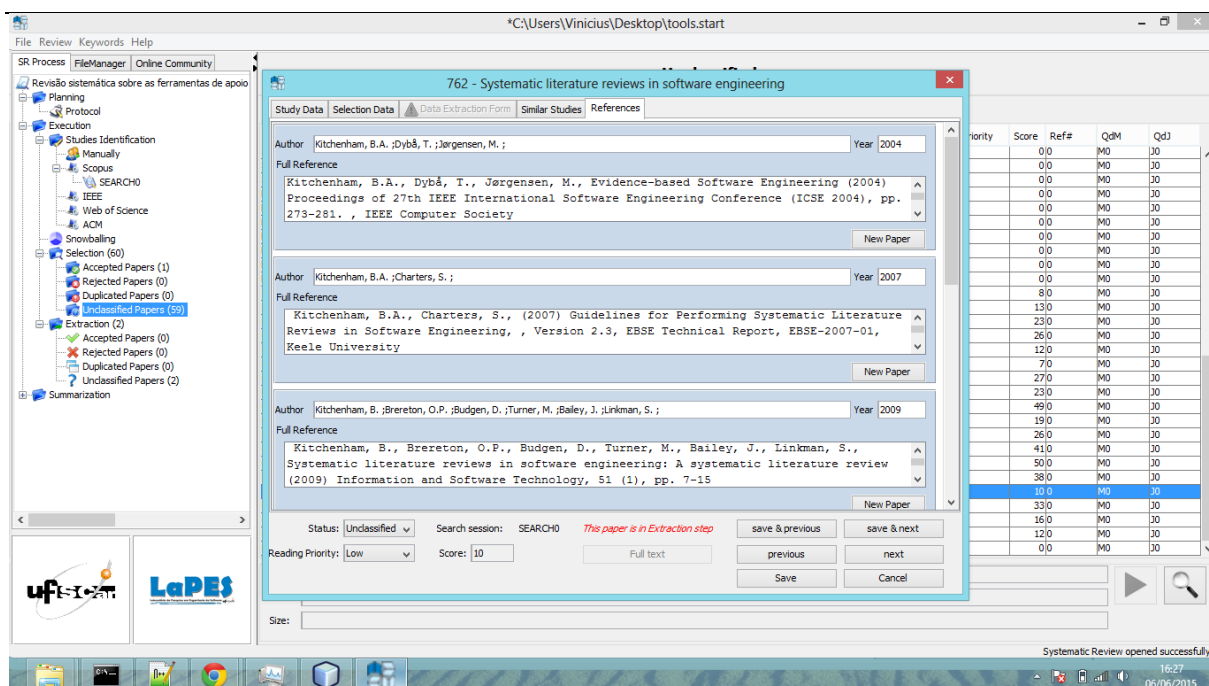


Figura 8. Aba de referências extraídas do *full text* pela ferramenta StArt

A Figura 9 mostra a interface da StArt para o pesquisador acompanhar os resultados encontrados por *snowballing* de forma conjunta. As referências são apresentadas em uma tabela e os critérios de priorização para cada referência encontrada são apresentados em forma de colunas. Pela interface, o pesquisador pode definir uma referência como relevante e um novo estudo é automaticamente criado e adicionado à pesquisa. O pesquisador tem a flexibilidade de escolher por qual dos critérios deseja ranquear as referências para a leitura. Caso o pesquisador tenha interesse em refinar a pesquisa, utilizando a mesma interface, ele terá acesso a recursos de frequência de conferências, periódicos e de autores recorrentes nas referências encontradas, permitindo identificar pesquisadores que trabalham no tema e os principais locais de publicação utilizados por eles.

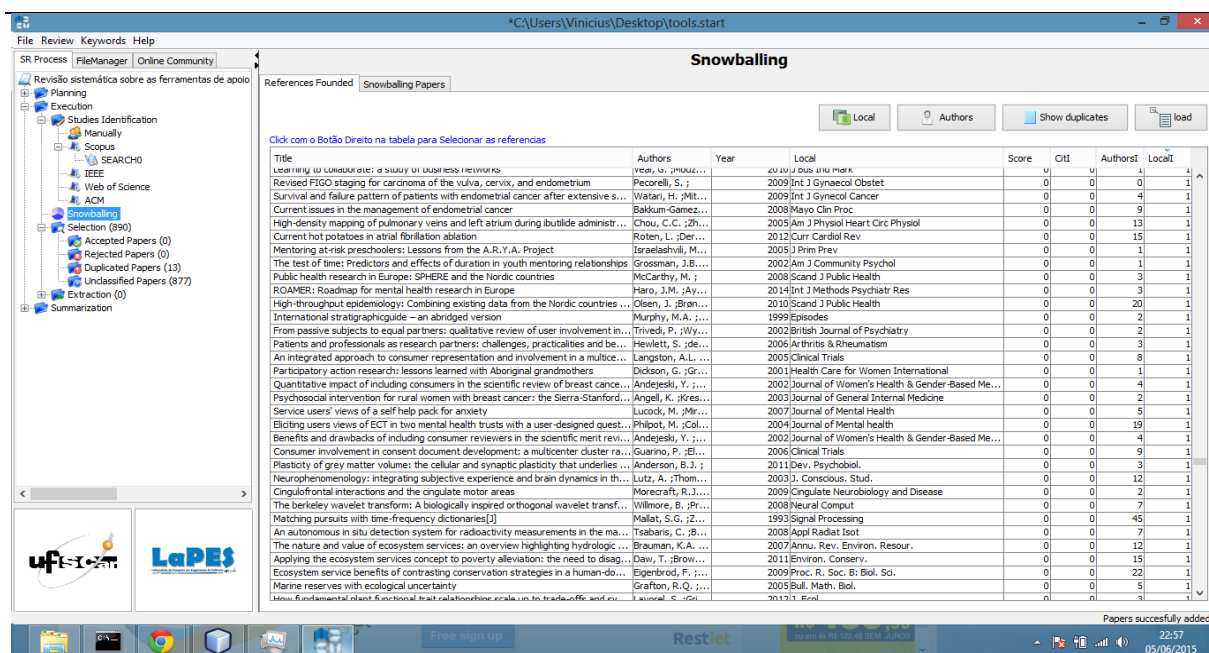


Figura 9. Interface para a execução de snowballing

Wohlin (2016) recomenda realizar *snowballing* utilizando o texto completo e avaliando os trechos de texto que representam uma citação. O pesquisador poderá fazer a avaliação durante a leitura de um estudo relevante. Foram incorporados recursos à ferramenta StArt para permitir a visualização de arquivos PDF e execução de buscas dentro dos arquivos. Assim o pesquisador poderá executar snowballing manualmente utilizando os recursos da ferramenta. Esses recursos ainda não estão disponíveis para a comunidade, mas deverão ser disponibilizados brevemente. A Figura 10 apresenta a interface da ferramenta com os recursos mencionados.

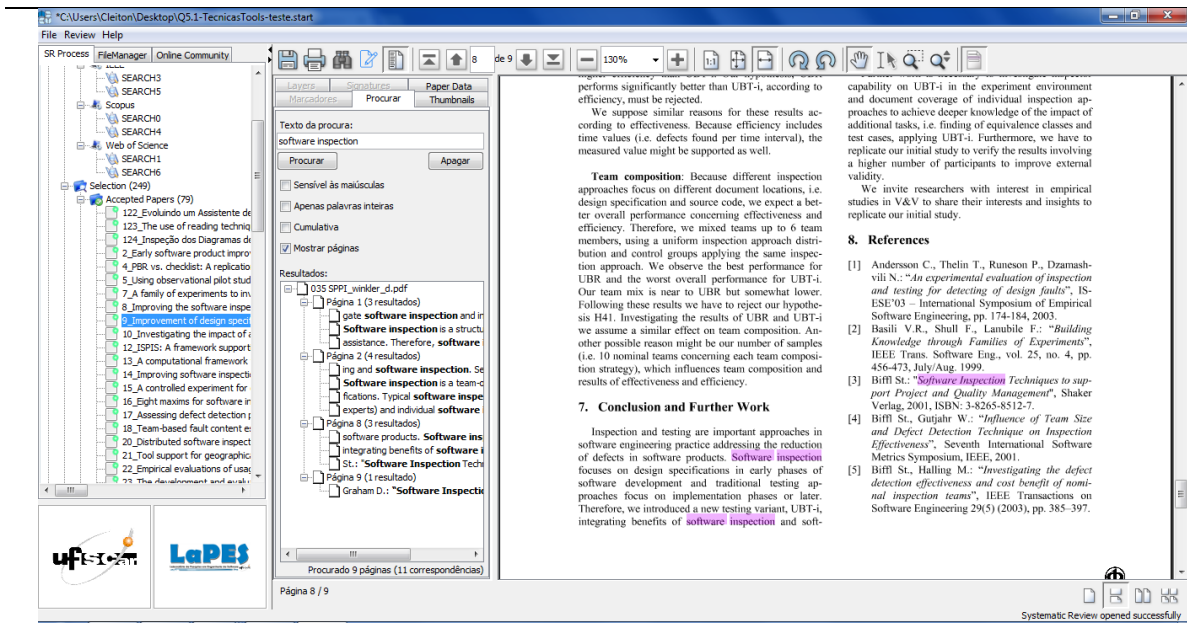


Figura 10. Interface da StArt para execução de *snowballing* manual com uso de PDF

Fabbri et al. (2012) apresentam o recurso de referências das referências da ferramenta StArt. Tanto o algoritmo utilizado para gerar a visualização quando o layout foram melhorados neste trabalho. A Figura 11 apresenta a visualização como proposta por Fabbri et al. (2012)

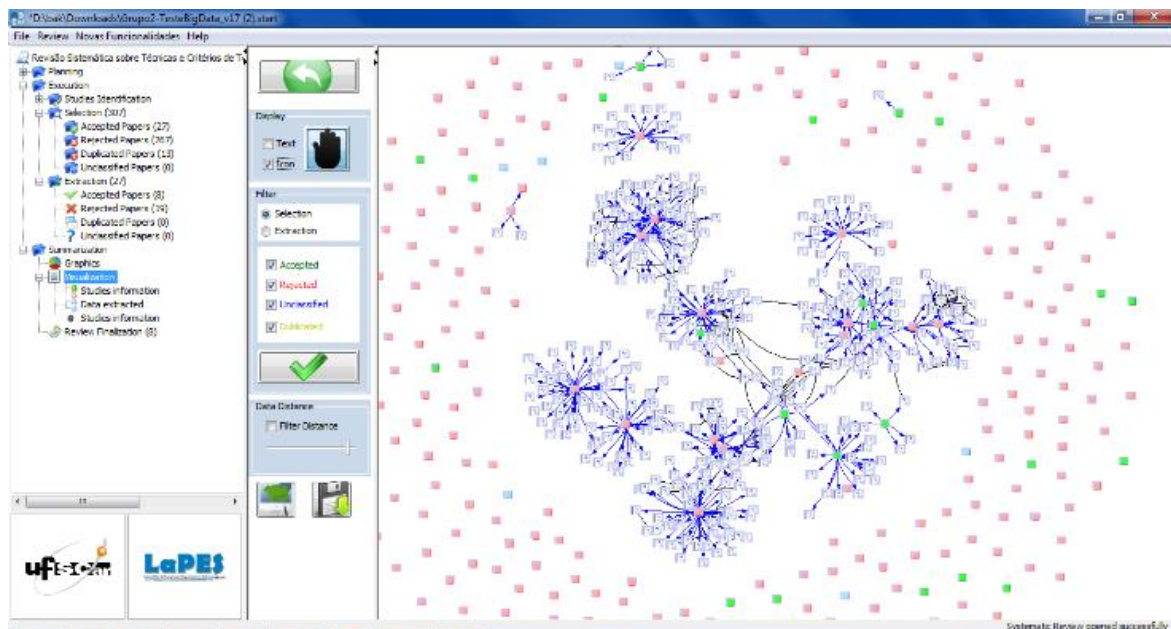


Figura 11. Grafo de referências das referências na ferramenta StArt (Fabbri et al., 2012)

Por meio da interface da Figura 11, o pesquisador tem acesso a uma metáfora visual para a execução de *snowballing*. Dentre os principais benefícios de utilizar a

interface, destaca-se a possibilidade de identificar agrupamentos (*clusters*) de forma visual. A Figura 12 apresenta na interface um exemplo de agrupamentos de uma pesquisa encontrados por meio de *snowballing*.

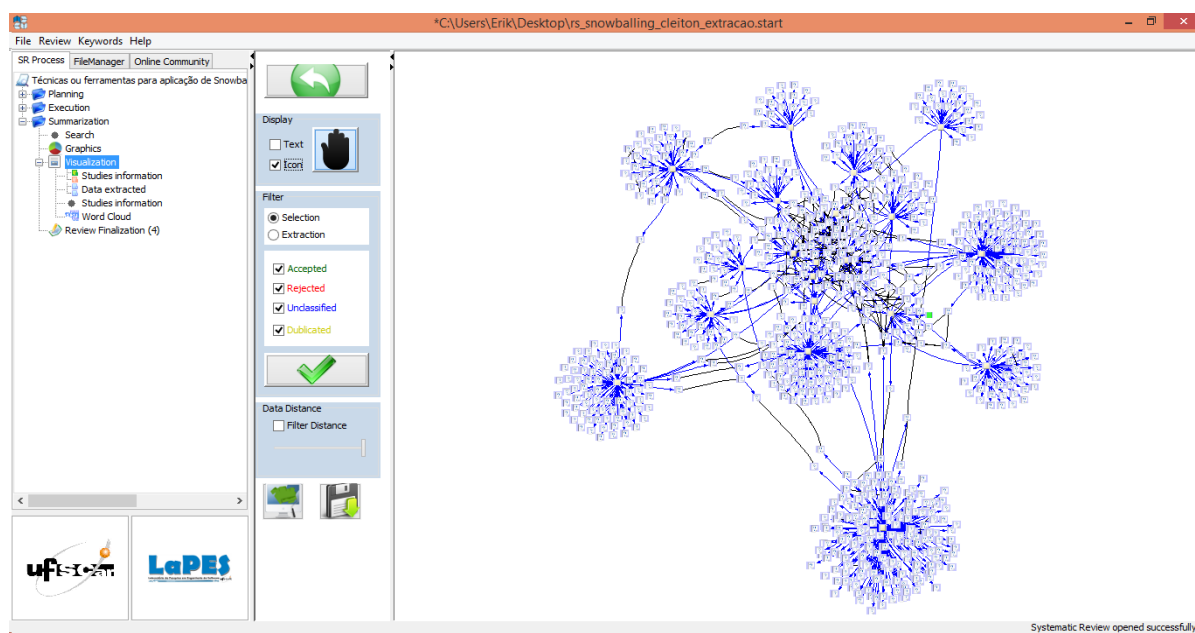


Figura 12. Interface para visualizar agrupamento

Uma vez aplicado o *snowballing*, o pesquisador poderá avaliar os estudos a fim de encontrar novos termos e palavras-chave para a *string* de busca. Existe a possibilidade dos novos estudos avaliados por *snowballing* utilizarem sinônimos ainda não contemplados na busca com *strings*. A Figura 13 apresenta a nuvem de palavras formadas na lista de referências. O pesquisador poderá gerar a nuvem para todas as referências ou apenas para os estudos relevantes selecionados para a pesquisa.

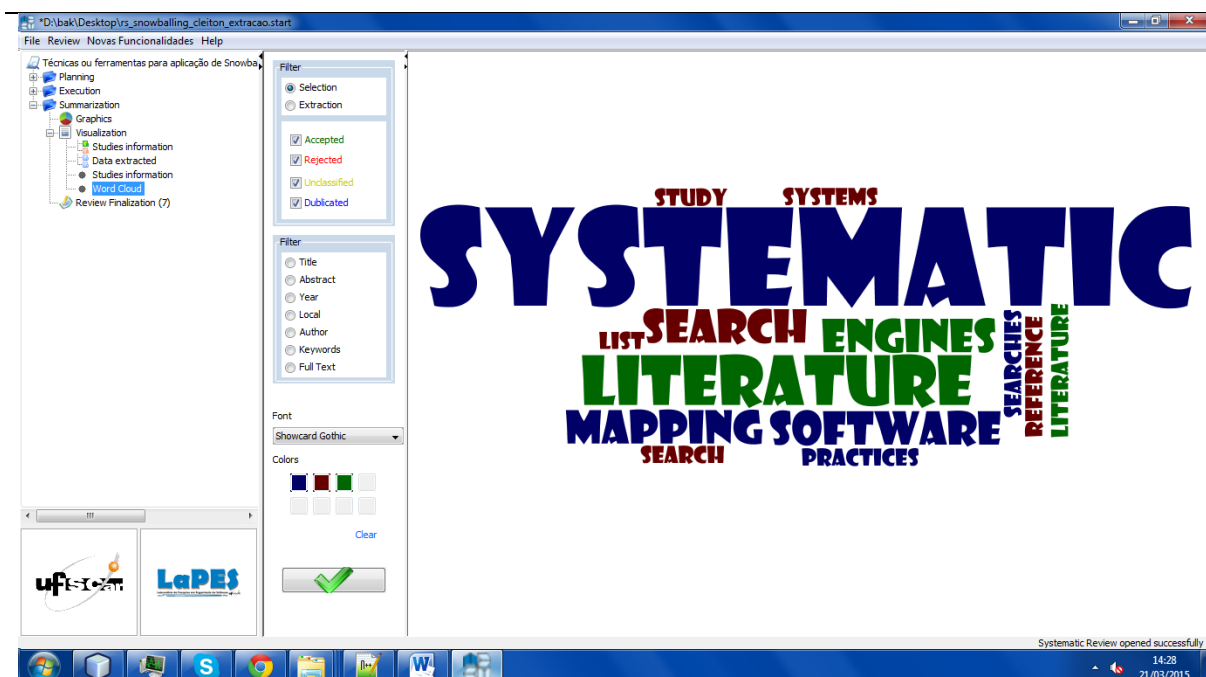


Figura 13. Interface para visualização de nuvens de palavras

4.8 Considerações Finais

Apesar de ser aceita como método de busca para estudos primários, a técnica *snowballing* apresenta grades desafios.

Neste capítulo foram apresentadas técnicas para a construção do conjunto inicial de estudos primários, critérios para a expansão de estudos e critérios de parada de novas buscas para as pesquisas. Uma estratégia completa para a aplicação de *snowballing* é definida como a combinação dos três critérios. Suporte computacional, pode ser utilizado para execução de *snowballing* em diferentes estratégias de aplicação. Existe uma série de desafios que dificultam a construção de um suporte computacional eficaz.

Para auxiliar na aplicação de *snowballing*, foram propostos critérios de priorização de estudos primários e recursos foram desenvolvidos na ferramenta StArt para auxiliar pesquisadores que desejam utilizar a técnica em seus estudos secundários. Tanto os critérios propostos quanto o suporte computacional foram malvos de estudos experimentais, que são apresentados no próximo capítulo.

Capítulo 5

ESTUDOS EXPERIMENTAIS

Este capítulo descreve os estudos experimentais realizados para a avaliação deste trabalho: um experimento para avaliar as dificuldades de aplicação de snowballing backward e forward, um estudo exploratório para avaliar os algoritmos de extração de referências e de cálculo de similaridade propostos, e um experimento para avaliar a aplicação dos critérios de priorização propostos neste trabalho.

5.1 Considerações Iniciais

A fim de melhor avaliar as estratégias para a aplicação de *snowballing* e o trabalho proposto, um conjunto de estudos experimentais foi conduzido. O primeiro trata-se de um experimento realizado para avaliar a aplicação de *snowballing backward* e *forward* em resultados de algumas pesquisas da área da computação, identificando as principais dificuldades reportadas pelos participantes com o intuito de auxiliar na elaboração dos critérios de priorização propostos neste trabalho. O segundo trata-se de um estudo exploratório para avaliar a eficácia dos algoritmos de extração de referências com base em arquivos BibTex e em textos completos dos estudos primários e de cálculo de similaridade propostos e incorporados na ferramenta StArt, os quais são partes importantes que tornam possível a utilização na prática dos critérios de priorização propostos neste trabalho. O terceiro trata-se de um experimento para avaliar a aplicação de *snowballing backward* nos resultados de uma busca com *strings* e avaliar ainda os critérios de priorização propostos, que são a principal contribuição deste trabalho. Os três estudos experimentais serão apresentados com mais detalhes nas próximas seções.

Algumas dificuldades são encontradas ao realizar experimentos que envolvem a execução de estudos secundários por participantes. As principais são a complexidade e a subjetividade existente ao conduzir estudos secundários, pois depende do nível de entendimento adquirido por cada participante, ainda que o

treinamento tenha sido o mesmo para todos os participantes, e da avaliação pessoal dos mesmos no que diz respeito à aceitação ou rejeição de estudos primários, onde a decisão final é estritamente pessoal, embora as orientações gerais dadas a todos os participantes fossem as mesmas.

Ressalta-se ainda que não foi possível conduzir estudos experimentais avaliando o uso de todas as estratégias de busca elencadas neste trabalho como conjunto inicial de estudos para aplicação de *snowballing* dada a complexidade para conduzir esse tipo de estudo. Segundo Badampudi, Wohlin e Petersen (2015) as duas formas de busca mais utilizadas para se conduzir estudos secundários em Engenharia de Software são: *snowballing* e buscas com string. Os estudos experimentais conduzidos avaliaram apenas a combinação dessas duas estratégias para a condução de estudos secundários, utilizando os resultados da busca com *string* como conjunto inicial de estudos do *snowballing*.

5.2 Experimento: um comparativo entre *snowballing backward* e *forward* e as principais dificuldades de execução

Para avaliar a melhoria dos resultados da busca e a efetividade de cada uma das técnicas de *snowballing* para encontrar novos resultados e determinar dificuldades para a aplicação de *snowballing*, um estudo experimental foi planejado e conduzido e os detalhes referentes ao planejamento, execução e resultados são descritos nessa seção.

5.2.1 Planejamento e Contexto do Experimento

Snowballing pode ser dividido em dois tipos: *snowballing backward* e *snowballing forward*. Esse experimento foi planejado com o objetivo principal de verificar se ambos os tipos de *snowballing* contribuem para a melhoria da atividade de busca com *strings*, auxiliando na identificação de estudos relevantes adicionais que não seriam recuperados apenas por meio de *strings* de busca. Além disso, como objetivo secundário, temos a identificação das principais dificuldades quando o pesquisador executa cada um dos tipos. O experimento foi planejado com o auxílio do processo GQM (BASILI; 1992). Sua descrição é apresentada na Tabela 10.

Tabela 10. Descrição do experimento planejado

Analisar	<i>Snowballing backward</i> e <i>forward</i> aplicada nos resultados de uma busca com <i>String</i> .
Com o propósito de	Avaliar a melhoria dos resultados da busca e a efetividade de cada uma das técnicas de snowballing para encontrar novos resultados e terminar as principais dificuldades para os participantes
Com respeito a	Precisão dos resultados encontrados
Do ponto de vista dos	Pesquisadores
No contexto de	Estudantes de pós-graduação em Computação da Universidade de Federal de São Carlos.

O experimento foi considerado uma das atividades da disciplina de pós-graduação de revisão sistemática do PPGCC da Universidade Federal de São Carlos. A disciplina recebeu inscrições apenas de alunos do PPGCC, diferentemente do que aconteceu no experimento anterior. *Snowballing* e estudos secundários faziam parte ementa da disciplina.

Wohlin *et al.* (2000) define quatro dimensões que caracterizam o contexto de um experimento. Esse é considerado do tipo *off-line* por não ser construído ou baseado na indústria, os participantes são estudantes de pós-graduação (mestrado ou doutorado), e o problema é um problema tipo *toy*, representando uma fração da pesquisa de cada participante.

Os resultados obtidos por esse estudo puderam auxiliar na melhoria da ferramenta StArt e a definir e refinar as estratégias propostas para *snowballing*. Também apresentaram caráter avaliativo das estratégias já definidas.

5.2.2 Seleção das Variáveis

As seguintes variáveis foram consideradas:

- A quantidade de participantes que encontraram estudos relevantes com *snowballing backward* ou *forward* separadamente e em conjunto.

- A quantidade de participantes que reportaram como dificuldade para executar *snowballing* cada uma das dificuldades previamente elencadas com base nos estudos bibliográficos.

5.2.3 Seleção dos Participantes

Conforme mencionado anteriormente, os participantes são alunos da disciplina de revisão sistemática do PPGCC da UFSCar. Eles não receberam nota pelo desempenho no experimento, apenas uma nota de participação na atividade. Por serem alunos de pós-graduação, os participantes já tiveram contato com pesquisa científica e continuaram a trabalhar com seus próprios tópicos de pesquisa durante o experimento. O número total de participantes foi 26.

Inicialmente eles foram contextualizados sobre o experimento e a pesquisa em andamento, e também sobre o projeto da ferramenta StArt. Foi-lhes explicado que, ao participar do experimento, estariam contribuindo com este trabalho de pesquisa. Todos os participantes se mostraram dispostos a participar e contribuir, além de considerarem a utilização da técnica avaliada em suas pesquisas futuramente.

5.2.4 Projeto Experimental

O experimento foi realizado de forma individual, e cada um dos participantes realizou o experimento em seu próprio tema de pesquisa. Apesar de trabalharem individualmente, para facilitar a compreensão, eles foram divididos em sete grupos de acordo com a área de pesquisa. A Tabela 11 mostra o número de participantes agrupados por área de pesquisa.

Tabela 11. Total de participantes por área de pesquisa

Grupo	Sigla	Nº de Participantes
Banco de dados	BD	2
Engenharia de Software	ES	7
Sistemas Distribuídos	SD	5
Inteligência Artificial	IA	6
Arquitetura e Redes de Computadores	AR	3
Processamento de Imagens	PI	1
Interação Humano Computador	IHC	2

O estudo foi executado em três dias. As atividades duraram quatro horas em cada dia.

- Dia 1: Os participantes definiram o tema de pesquisa e construíram o protocolo de suas revisões em comum acordo com seus orientadores de pesquisa.
- Dia 2: Os participantes construíram as *strings* de busca e refinaram a busca para retornar o melhor resultado possível. Os participantes realizaram a classificação de todos os estudos primários identificados ao executar a *string* de busca e empacotaram os dados na ferramenta StArt.
- Dia 3: Os participantes aplicaram as técnicas de *snowballing backward* e *snowballing forward* manualmente com o auxílio do Microsoft Excel. Os resultados foram coletados por meio de planilhas eletrônicas geradas a partir da ferramenta StArt para posterior análise.

5.2.5 Instrumentação

Para realização do experimento foram utilizadas a ferramenta StArt para o empacotamento do protocolo, da *string* de busca e dos dados de seleção dos estudos. A máquina de busca SCOPUS foi utilizada para a realização da busca com *strings*. O Microsoft Excel foi utilizado para a catalogação dos dados de execução de *snowballing backward* e *forward*. O Google Forms foi utilizado para preenchimento dos questionários.

5.2.6 Ameaças à Validade

Validade Interna:

5.2.7 Interferência no desempenho do participante devido à atribuição de nota de participação na atividade: como as notas foram atribuídas pela participação do estudante, considera-se que o risco associado ao desempenho tenha sido minimizado.

5.2.8 Erro no registro dos dados coletados: esse risco não pode ser minimizado pois cada participante foi responsável pela execução das buscas e importação na ferramenta. Os participantes deveriam salvar os dados ao final das etapas de execução do estudo.

5.2.9 Uma possível ameaça foram as buscas realizadas na Internet por cada participante para localizar referências por *snowballing*. Os resultados poderiam não ter todos os dados referentes às listas de referências e de citações indexadas.

Validade Externa:

1. Os resultados não podem ser generalizados para outros perfis de pesquisas. Essa ameaça não pode ser minimizada, pois não houve possibilidade de aplicar o experimento com outros tipos de pesquisadores e com diferentes níveis de experiência.

5.2.10 Os grupos foram determinados em acordo com a área de pesquisa, e a quantidade de estudos dependeu dos resultados encontrados na busca. É possível que, ao utilizar outras bases de dados ou mudar as buscas, o resultado seja diferente.

5.2.11 A execução do estudo em momentos futuros implicaria em resultados diferentes na execução da busca nas máquinas de busca e, conseqüentemente, a amostra de estudos avaliados seria diferente.

Validade de Conclusão:

1. A ausência da aplicação de testes estatísticos pode ser considerada uma ameaça de conclusão para os resultados desse estudo. Apenas as medias aritméticas foram aplicadas.

5.2.7 Método

As seguintes tarefas foram realizadas no experimento:

- Tarefa 1: preencher um questionário de caracterização, o qual incluía informações sobre detalhes pessoais e técnicos dos participantes, incluído o tema de pesquisa (dia 1).
- Tarefa 2: Todos os participantes receberam um treinamento sobre RS e sobre a ferramenta StArt. Esse treinamento visou ao nivelamento técnico dos participantes (dia 1).

-
- Tarefa 3: Todos os participantes construíram o protocolo completo para a revisão sistemática em seu tema de pesquisa (dia 1).
 - Tarefa 4: Todos os participantes receberam treinamento sobre buscas com *strings* e sobre a máquina de busca SCOPUS (dia 2).
 - Tarefa 5: Os participantes construíram as *strings* de busca e realizaram a busca na SCOPUS. Os resultados foram importados na ferramenta StArt. Os participantes realizaram a primeira seleção de todos os estudos recuperados por meio da busca com *strings* (dia 2).
 - Tarefa 6: Os participantes receberam um treinamento prévio sobre a técnica *snowballing*, aplicarem-na manualmente em uma amostra de estudos. A amostra foi escolhida com base nos estudos pertencentes ao primeiro e quarto quadrantes do algoritmo SCAS, descrito na seção 2.5 do Capítulo 2. O Algoritmo foi escolhido para permitir selecionar uma amostra heterogênea de estudos aceitos e rejeitados. Depois disso, foram apresentados aos participantes recursos de *snowballing* implementados na ferramenta StArt (dia 3).
 - Tarefa 7: Os participantes foram instruídos a realizar uma iteração de *snowballing forward* e *backward* manualmente em todos os estudos aceitos da amostra selecionada. Os participantes relataram os resultados em uma planilha do Microsoft Excel (dia 3).

5.2.8 Coleta de Dados e Disponibilização dos dados

Os dados foram coletados por meio de questionários preenchidos por cada participante do experimento. Além dos questionários, cada participante entregou o arquivo com extensão .start contendo os dados coletados para análise posterior e uma planilha Excel com os dados resultantes da aplicação de *snowballing*.

Os dados estão acessíveis em formato de planilhas eletrônicas em: <http://goo.gl/Sp3xyu>

5.2.9 Análise e Interpretação

As análises foram executadas com o uso da ferramenta Excel e os resultados são apresentados a seguir, divididos em análise descritiva e estatística.

5.2.9.1 Análise Descritiva

A Tabela 12 apresenta os dados sumarizados quantitativamente dos participantes do experimento. A primeira linha de dados, isto é, a linha seguinte a de cabeçalho da tabela, mostra o total de participantes por grupo. A segunda e terceira linhas mostram o número de participantes que encontraram novos estudos relevantes aplicando *snowballing backward* e *forward*, respectivamente. A quarta linha indica o número de participantes que não encontraram novos estudos aplicando *snowballing*. A quinta, sexta, sétima e oitava linhas da tabela referem-se às dificuldades encontradas pelos participantes e, por fim, as duas últimas linhas mostram quantos participantes consideram *snowballing backward* mais trabalhoso que *snowballing forward* e vice-versa.

Partindo dos resultados apresentados na Tabela 12, observa-se na terceira e quarta linha de dados que todos os participantes encontraram novas referências relevantes ao aplicar tanto *snowballing backward* quanto *snowballing forward*. Com relação às dificuldades encontradas para aplicação de *snowballing*, observa-se nas linhas 4 a 7 que, um total de 15 participantes apontam como dificuldade localizar as referências relevantes (58%), 16 participantes apontam como dificuldade identificar as duplicidades (62%), 13 participantes apontam como dificuldade avaliar as citações apenas por título, autor e conferência ou periódico (50%) e 11 participantes apontam como dificuldade avaliar a qualidade dos estudos localizados (42%). A Tabela 12 também apresenta os valores divididos por áreas de pesquisa agrupadas (grupo).

Com relação às duas últimas linhas da tabela, 14 participantes relataram que o maior esforço foi empregado para realizar *snowballing backward* (54%), cinco participantes apontaram que *snowballing forward* demanda mais esforço (19%) para sua execução e, para os 27 participantes restantes, as duas técnicas demandam o mesmo esforço. Ressalta-se que os valores de porcentagem estão arredondados em duas casas decimais.

Tabela 12. Dados quantitativos referentes à aplicação de *snowballing* e dificuldades encontradas

Descrição	GRUPO							Total
	BD	ES	SD	IA	AR	PI	IHC	
Nº de participantes	2	7	5	6	3	1	2	26
Nº de participantes que encontraram novos estudos aplicando <i>snowballing backward</i>	2	7	5	6	3	1	2	26
Nº de participantes que encontraram novos estudos aplicando <i>snowballing forward</i>	2	7	5	6	3	1	2	26
Nº de participantes que não encontraram novos estudos aplicando <i>snowballing</i>	0	0	0	0	0	0	0	0
Nº de participantes que apontam como dificuldade localizar as referências relevantes	1	5	3	3	2	1	0	15
Nº de participantes que apontam como dificuldade identificar as duplicidades	2	4	2	4	2	0	2	16
Nº de participantes que apontam como dificuldade avaliar as citações apenas por título, autor e conferência.	1	4	3	4	1	0	1	13
Nº de participantes que apontam como dificuldade avaliar a qualidade dos estudos localizados	1	2	3	0	2	1	2	11
Nº de participantes que apontam <i>snowballing backward</i> como mais trabalhoso que <i>snowballing forward</i> .	1	2	3	3	2	1	2	14
Nº de participantes que apontam <i>snowballing forward</i> como mais trabalhoso que <i>snowballing backward</i> .	0	3	1	0	1	0	0	5

A Figura 14 apresenta a avaliação dos participantes em relação ao esforço exigido para a aplicação dos dois tipos de *snowballing*. A parte cinza representa o número de participantes que considera que *snowballing backward* exige maior esforço que *snowballing forward* (14 participantes - 54%). A parte laranja representa o contrário (5 participantes - 19%). Por fim, a parte azul indica o número de participantes que consideram equivalente o nível de esforço exigido para aplicação dos dois tipos de *snowballing* (7 participantes - 27%).

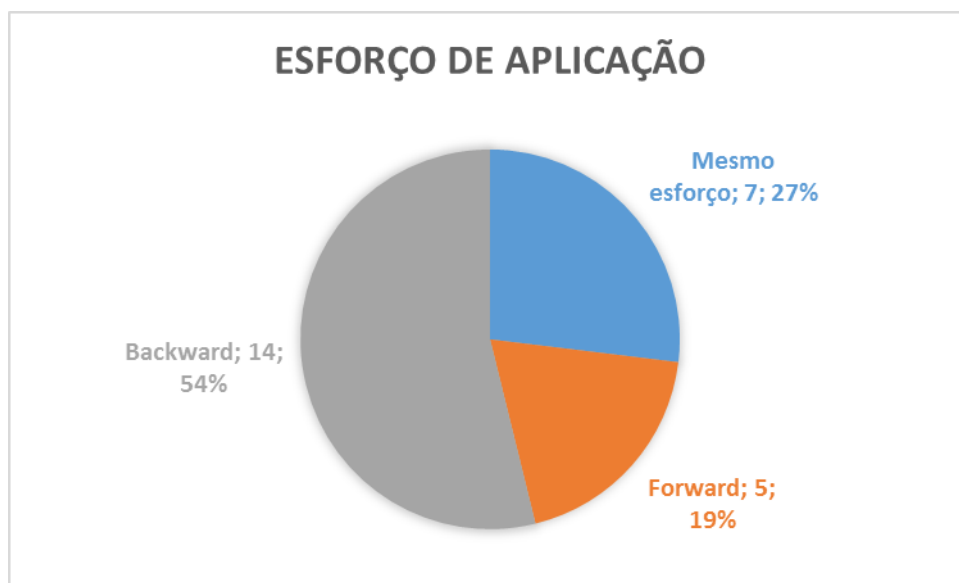


Figura 14. Avaliação de esforço dos participantes ao aplicar *snowballing*

Com relação à melhoria de resultados, a Tabela 13 apresenta as quantidades de novos estudos encontrados ao aplicar as técnicas de busca por *string* e, posteriormente, ao aplicar *snowballing forward* e *snowballing backward* sobre o conjunto resultante da busca por *string*. A primeira coluna apresenta a identificação definida para o participante, a segunda coluna apresenta o grupo definido, na terceira coluna são apresentados o número de estudos relevantes identificados pela busca com *strings*, na quarta coluna é apresentado o número de estudos relevantes encontrados por *snowballing forward*. A última coluna apresenta o número de estudos relevantes encontrados por *snowballing backward*. Com base nos dados da tabela, ao observarmos as linhas (que representam os participantes) e as duas últimas colunas (que representam os resultados de *snowballing*), podemos observar que *snowballing* proporcionou melhoras nos resultados para todos os participantes envolvidos. É posto que não é possível inferir conclusões sobre a qualidade e a relevância dos novos estudos encontrados.

Tabela 13. Dados quantitativos referentes à aceitação de estudos pelas técnicas

Participante	Grupo	Estudos Aceitos (string de busca)	Estudos Aceitos (snowballing forward)	Estudos Aceitos (snowballing backward)
1	BD	70	4	10
2	ES	53	23	33
3	ES	18	15	1
4	ES	75	31	95
5	SD	160	2	10
6	IA	130	10	12
7	AR	60	29	40
8	IA	119	2	1
9	PI	36	15	30
10	ES	117	48	71
11	AR	21	71	37
12	SD	38	9	23
13	IHC	208	10	15
14	IHC	33	4	11
15	SD	28	31	13
16	ES	97	73	7
17	AR	155	27	53
18	ES	60	38	79
19	IA	148	15	8
20	BD	13	3	8
21	IA	134	4	39
22	SD	121	7	16
23	SD	65	8	19
24	IA	380	9	10
25	IA	10	3	3
26	ES	133	34	34

5.2.9.2 Análise Estatística

A Tabela 14 apresenta a porcentagem de estudos aceitos por meio da aplicação de *string* de busca, de *snowballing forward*, de *snowballing backward* e ainda da aplicação em conjunto de *snowballing forward* e *backward* por participante. A primeira coluna apresenta a identificação do participante, a segunda coluna apresenta o grupo, a terceira coluna apresenta o número total de estudos relevantes identificados com a aplicação conjunta de *snowballing backward* e *forward*. A quarta, quinta, sexta e sétima colunas representam, respectivamente, os valores aproximados em porcentagem de estudos relevantes identificados por cada técnica (*string* de busca, *snowballing forward*, *snowballing backward*, e *snowballing forward* e *backward* combinados).

Tabela 14. Porcentagem de estudos aceitos encontrados por meio de cada técnica

Participante	Grupo	Total	<i>String</i>	<i>Forward</i>	<i>Backward</i>	<i>Forward + Backward</i>
1	BD	84	83%	5%	12%	17%
2	ES	109	49%	21%	30%	51%
3	ES	34	53%	44%	3%	47%
4	ES	201	37%	15%	47%	63%
5	SD	172	93%	1%	6%	7%
6	IA	152	86%	7%	8%	14%
7	AR	129	47%	22%	31%	53%
8	IA	122	98%	2%	1%	2%
9	PI	81	44%	19%	37%	56%
10	ES	236	50%	20%	30%	50%
11	AR	129	16%	55%	29%	84%
12	SD	70	54%	13%	33%	46%
13	IHC	233	89%	4%	6%	11%
14	IHC	48	69%	8%	23%	31%
15	SD	72	39%	43%	18%	61%
16	ES	177	55%	41%	4%	45%
17	AR	235	66%	11%	23%	34%
18	ES	177	34%	21%	45%	66%
19	IA	171	87%	9%	5%	13%
20	BD	24	54%	13%	33%	46%
21	IA	177	76%	2%	22%	24%
22	SD	144	84%	5%	11%	16%
23	SD	92	71%	9%	21%	29%
24	IA	399	95%	2%	3%	5%
25	IA	16	63%	19%	19%	38%
26	ES	201	66%	17%	17%	34%

As médias aritméticas foram calculadas para os valores exibidos na Tabela 14 e são apresentadas na Tabela 15. Os valores das médias foram calculados para buscas com *string* (primeira coluna), *snowballing backward* (segunda coluna), *snowballing forward* (terceira coluna) e ambos (quarta coluna). A média aritmética das pesquisas foi escolhida para dar o mesmo peso de relevância para cada pesquisa, independentemente do número de estudos encontrados. A média geral seria uma alternativa, contudo, como as pesquisas possuíam tamanhos diferentes, esse tipo de média não seria indicada nesse contexto, pois pesquisas muito grandes teriam maior influência nos resultados.

Tabela 15. Médias individuais de porcentagens de estudos identificados por método de busca

	STRING	FORWARD	BACKWARD	FORWARD+BACKWARD
Média Individual	64%	16%	20%	36%

Enfim, com base nas análises realizadas, é possível dizer que ambos os tipos de *snowballing* melhoraram os resultados de uma busca com *strings*, como esperado. As melhorias são apresentadas na Tabela 14 com os resultados individuais dos participantes e na Tabela 15 com as médias aritméticas calculadas para cada tipo. Além disso, com base nos dados da Tabela 12 e na Figura 15, pode-se dizer que os participantes encontraram dificuldades na aplicação das técnicas. Com relação ao impacto de cada dificuldade, no contexto desse experimento, a maior dificuldade elencada foi identificar as duplicações dos estudos encontrados.

5.3 Estudo exploratório: análise da eficácia das técnicas de extração de referências e dos algoritmos de similaridade ao aplicar *snowballing* com suporte da ferramenta StArt.

Foram incorporados à ferramenta StArt recursos para a execução de *snowballing*. Os algoritmos implementados na StArt para permitir o suporte à aplicação da técnica não são precisos, e as dificuldades são descritas na Seção 3.7 do Capítulo 3. Pode-se dividir a execução de *snowballing* em duas tarefas principais: a coleta das referências e a classificação das referências encontradas. Assim, este estudo de caso diz respeito diretamente a essas tarefas, tendo o objetivo de analisar as técnicas de coleta de referências e os algoritmos de cálculo de similaridade implementados na ferramenta StArt para reconhecimento de padrões de textos, os quais permitem a aplicação de *snowballing*.

5.3.1 Seleção do Objeto de Estudo

Para atingir o objetivo, planejou-se a utilização de um estudo secundário publicado (Hernades, Belgamo, Fabbri; 2013), que é um MS. Ele foi escolhido por se tratar de um estudo publicado e, assim sendo, ter sido avaliado por revisores e representar um caso real de pesquisa. Além disso, os autores do estudo fazem parte do mesmo grupo de pesquisa do autor deste trabalho, o que facilita a comunicação com os autores. Por fim, os autores do MS disponibilizaram o arquivo no formato

.start utilizado pela ferramenta StArt, o qual possuía todos os dados empacotados referentes ao MS, facilitando a execução na ferramenta e análises posteriores.

No estudo selecionado, foram testadas as duas técnicas de reconhecimento de referências implementadas na StArt: reconhecimento pelo texto completo (arquivo no formato PDF) e pelo arquivo de referências (no formato BibTex). Além disso, foram avaliados os algoritmos de similaridade utilizados para comparar referências e determinar duplicações e o número de citações internas do estudo secundário.

5.3.2 Seleção das Variáveis

As seguintes variáveis foram consideradas na realização deste estudo exploratório:

- Número de referências identificadas por estudo pela técnica de extração do texto completo.
- Número de referências identificadas por estudo pela técnica de *parser* do arquivo de referências.
- Número de referências duplicadas encontradas nos resultados da técnica de extração do texto completo
- Número de referências duplicadas encontradas nos resultados da técnica de *parser* do arquivo de referências.

5.3.3 Método

Para a realização do estudo, o arquivo .start do MS foi considerado. O arquivo existente era de uma versão anterior da ferramenta StArt que não possuía suporte a referências. Assim sendo, as mesmas tiveram de ser adicionadas ao MS para permitir a realização do estudo de caso.

Para adicionar as referências, primeiramente todos os arquivos PDF contendo o texto completo foram recuperados e posteriormente carregados na ferramenta StArt. Depois disso, todas as buscas descritas foram reexecutadas nas máquinas de busca e importadas novamente na ferramenta. Isso fez-se necessário pois os arquivos BibTex originais não estavam disponíveis e era necessário obtê-los. Todos os novos estudos encontrados, isto é, aqueles adicionais aos existentes

inicialmente, forma rejeitados, ficando apenas com o conjunto original de estudos aceitos MS publicado.

Uma vez que essas atividades foram realizadas, todas as referências de *snowballing backward* obtidas pelos algoritmos da StArt foram catalogadas e estavam disponíveis na ferramenta. Assim, os algoritmos para comparação de similaridade das referências foram aplicados e os dados do MS foram exportados para uma planilha Excel para serem analisados.

5.3.4 Instrumentação

Para a execução deste estudo foi utilizada a Ferramenta StArt com os novos recursos desenvolvidos neste trabalho, o Microsoft Excel para catalogação dos dados e construção de gráficos e tabelas, e o Software R para análises estatísticas.

5.3.5 Questões a serem respondidas

As seguintes questões foram consideradas para o estudo exploratório:

- A combinação das técnicas de extração de referências por meio do texto completo e por meio do *parser* do arquivo de referências BibTex é mais eficaz que utilizar as técnicas separadamente?
- O novo algoritmo de similaridade proposto é mais eficaz que o algoritmo anterior de Salton e Allan (1994) apresentado por Fabbri *et al.* (2012) para comparar referências?

5.3.6 Ameaças à Validade

- **Validade Interna:** os dados tiveram pouca intervenção manual durante o estudo de caso. Contudo, alguns dados tiveram de ser removidos manualmente para possibilitar o isolamento de erros entre os algoritmos de recuperação de referências e os algoritmos para identificar duplicações. Nessas etapas é possível que existam erros por conta do procedimento manual realizado.
- **Validade de Conclusão:** para a técnica de extração dos dados por meio de arquivo de referências, erros ou falta de completude da lista de

referências exportadas pela base de dados Scopus são ameaças à validade de conclusão. Como a avaliação da eficácia da técnica depende de dados exportados pela máquina de busca, torna-se mais difícil avaliar o percentual de erro existente apenas por conta da aplicação da técnica.

- **Validade Externa:** o uso de apenas um MS no estudo de caso faz com que os dados não possam ser generalizados e fiquem restritos somente ao contexto do estudo secundário utilizado. Para minimizar esse problema, mais de um conjunto de estudos secundários representando casos reais deveria ser avaliado. Tal procedimento é mencionado pelo autor no Capítulo 6 como trabalho futuro.

5.3.7 Disponibilização dos Dados Coletados

Os dados foram coletados utilizando a ferramenta StArt e o Microsoft Excel e estão acessíveis em: <http://goo.gl/jCRpSM>

5.3.8 Análise e Interpretação

Os resultados coletados e as análises são apresentados nesta seção em dois tópicos: análise descritiva e análise estatística.

5.3.8.1 Análise Descritiva

O conjunto de estudos aceitos por Hernandez, Belgamo e Fabbri (2013) é composto por 79 estudos primários e pode ser visto na Tabela 37 do Apêndice D.

A Tabela 16 apresenta a quantidade de referências identificadas por cada técnica de extração de referências (PDF ou BibTex) implementadas na StArt, aplicadas isoladamente ou em conjunto e comparadas à quantidade real de referências existentes para cada estudo. A primeira coluna é a identificação do estudo, seguindo a numeração definida na Tabela 37. A segunda coluna mostra a quantidade real de referências do estudo e as demais mostram as quantidades de referências identificadas e perdidas com a aplicação das técnicas isoladamente (colunas 3 a 6) ou em conjunto (colunas 7 e 8). A soma por linha, das colunas 3 e 4,

5 e 6, e 7 e 8 deve sempre ser igual ao valor na linha da coluna 2, pois a soma de todas as referências identificadas com as referências perdidas deve sempre ser igual ao valor total de referências reais em cada estudo.

Tabela 16. Relação de referências encontradas e erro obtido

Quantidade de Referências							
#	Real	Identificadas (PDF)	Perdidas (PDF)	Identificadas (BibTex)	Perdidas (BibTex)	Identificadas (BibTex+PDF)	Perdidas (BibTex+PDF)
1	18	18	0	14	4	18	0
2	32	30	2	27	5	30	2
3	21	19	2	17	4	19	2
4	46	10	36	46	0	46	0
5	18	12	6	18	0	18	0
6	26	25	1	25	1	25	1
7	29	27	2	28	1	28	1
8	12	11	1	12	0	12	0
9	22	22	0	22	0	22	0
10	20	20	0	18	2	20	0
11	23	23	0	22	1	23	0
12	20	13	7	18	2	18	2
13	34	32	2	34	0	34	0
14	20	19	1	20	0	20	0
15	16	14	2	14	2	14	2
16	49	19	30	48	1	48	1
17	3	0	3	3	0	3	0
18	20	20	1	20	0	20	0
19	18	18	0	18	0	18	0
20	38	38	0	37	1	38	0
21	21	21	0	21	0	21	0
22	15	14	1	15	0	15	0
23	36	26	10	0	36	26	10
24	27	25	2	27	0	27	0
25	23	20	3	23	0	24	0
26	18	17	1	0	18	17	1
27	22	21	1	22	0	22	0
28	18	18	0	0	18	18	0
29	21	21	0	0	21	21	0
30	27	24	3	27	0	27	0
31	27	25	2	27	0	27	0
32	8	0	8	8	0	8	0
33	23	16	7	21	2	21	2
34	20	14	6	0	20	14	6
35	20	20	0	0	20	20	0

36	26	0	26	0	26	0	26
37	23	23	0	22	1	23	0
38	32	31	1	0	32	31	1
39	15	14	1	0	15	14	1
40	20	20	0	0	20	20	0
41	18	12	6	0	18	12	6
42	20	14	6	0	20	14	6
43	53	51	2	0	53	51	2
44	13	12	1	0	13	12	1
45	25	21	4	0	25	21	4
46	15	15	0	0	15	15	0
47	30	21	9	0	30	21	9
48	15	10	5	0	15	10	5
49	11	10	1	0	11	10	1
50	20	11	9	0	20	11	9
51	23	7	16	0	23	7	16
52	23	21	2	0	23	21	2
53	57	55	2	0	57	55	2
54	20	0	20	0	20	0	20
55	22	20	2	0	22	20	2
56	34	30	4	0	34	30	4
57	19	18	1	0	19	18	1
58	19	19	0	5	14	19	0
59	22	18	4	0	22	18	4
60	18	0	18	0	18	0	18
61	18	0	18	0	18	0	18
62	31	31	0	0	31	31	0
63	19	19	0	0	19	19	0
64	23	20	3	19	4	20	3
65	14	13	1	14	0	14	0
66	16	15	1	0	16	15	1
67	8	7	1	7	1	7	1
68	36	24	12	0	36	24	12
69	4	4	0	0	4	4	0
70	39	36	3	0	39	36	3
71	36	26	10	31	5	31	5
72	23	23	0	0	23	23	0
73	37	4	33	0	37	4	33
74	26	26	0	0	26	26	0
75	27	20	7	27	0	27	0
76	10	10	0	0	10	10	0
77	19	19	0	0	19	19	0

78	21	20	1	0	21	20	1
79	12	12	0	0	12	12	0

A Tabela 17 apresenta os valores de erros e acertos (em porcentagem) de identificação das referências com a aplicação das técnicas de extração de referências quando comparadas à quantidade real de referências existentes por estudo. Os acertos e erros foram calculados com base nas informações apresentadas na Tabela 16. O percentual de acerto e erro é calculado dividindo-se a quantidade de referências identificadas e perdidas, respectivamente, pela quantidade real de referências para cada estudo conforme dados exibidos na Tabela 16. Ao somar o percentual por linha, das colunas de erro e acerto para cada técnica, o valor deve ser 100%.

Tabela 17. Porcentagem de erro e acerto das estratégias

Nº	% Acerto (PDF)	% Erro (PDF)	% Acerto (BibTex)	% Erro (BibTex)	% Acerto (BibTex+PDF)	% Erro (BibTex+PDF)
1	100%	0%	78%	22%	100%	0%
2	94%	6%	84%	16%	94%	6%
3	90%	10%	81%	19%	90%	10%
4	22%	78%	100%	0%	100%	0%
5	67%	33%	100%	0%	100%	0%
6	96%	4%	96%	4%	96%	4%
7	93%	7%	97%	3%	97%	3%
8	92%	8%	100%	0%	100%	0%
9	100%	0%	100%	0%	100%	0%
10	100%	0%	90%	10%	100%	0%
11	100%	0%	96%	4%	100%	0%
12	65%	35%	90%	10%	90%	10%
13	94%	6%	100%	0%	100%	0%
14	95%	5%	100%	0%	100%	0%
15	88%	13%	88%	13%	88%	13%
16	39%	61%	98%	2%	98%	2%
17	0%	100%	100%	0%	100%	0%
18	100%	5%	100%	0%	100%	0%
19	100%	0%	100%	0%	100%	0%
20	100%	0%	97%	3%	100%	0%
21	100%	0%	100%	0%	100%	0%
22	93%	7%	100%	0%	100%	0%
23	72%	28%	0%	100%	72%	28%
24	93%	7%	100%	0%	100%	0%
25	87%	13%	100%	0%	100%	-0%

26	94%	6%	0%	100%	94%	6%
27	95%	5%	100%	0%	100%	0%
28	100%	0%	0%	100%	100%	0%
29	100%	0%	0%	100%	100%	0%
30	89%	11%	100%	0%	100%	0%
31	93%	7%	100%	0%	100%	0%
32	0%	100%	100%	0%	100%	0%
33	70%	30%	91%	9%	91%	9%
34	70%	30%	0%	100%	70%	30%
35	100%	0%	0%	100%	100%	0%
36	0%	100%	0%	100%	0%	100%
37	100%	0%	96%	4%	100%	0%
38	97%	3%	0%	100%	97%	3%
39	93%	7%	0%	100%	93%	7%
40	100%	0%	0%	100%	100%	0%
41	67%	33%	0%	100%	67%	33%
42	70%	30%	0%	100%	70%	30%
43	96%	4%	0%	100%	96%	4%
44	92%	8%	0%	100%	92%	8%
45	84%	16%	0%	100%	84%	16%
46	100%	0%	0%	100%	100%	0%
47	70%	30%	0%	100%	70%	30%
48	67%	33%	0%	100%	67%	33%
49	91%	9%	0%	100%	91%	9%
50	55%	45%	0%	100%	55%	45%
51	30%	70%	0%	100%	30%	70%
52	91%	9%	0%	100%	91%	9%
53	96%	4%	0%	100%	96%	4%
54	0%	100%	0%	100%	0%	100%
55	91%	9%	0%	100%	91%	9%
56	88%	12%	0%	100%	88%	12%
57	95%	5%	0%	100%	95%	5%
58	100%	0%	26%	74%	100%	0%
59	82%	18%	0%	100%	82%	18%
60	0%	100%	0%	100%	0%	100%
61	0%	100%	0%	100%	0%	100%
62	100%	0%	0%	100%	100%	0%
63	100%	0%	0%	100%	100%	0%
64	87%	13%	83%	17%	87%	13%
65	93%	7%	100%	0%	100%	0%
66	94%	6%	0%	100%	94%	6%
67	88%	13%	88%	13%	88%	13%

68	67%	33%	0%	100%	67%	33%
69	100%	0%	0%	100%	100%	0%
70	92%	8%	0%	100%	92%	8%
71	72%	28%	86%	14%	86%	14%
72	100%	0%	0%	100%	100%	0%
73	11%	89%	0%	100%	11%	89%
74	100%	0%	0%	100%	100%	0%
75	74%	26%	100%	0%	100%	0%
76	100%	0%	0%	100%	100%	0%
77	100%	0%	0%	100%	100%	0%
78	95%	5%	0%	100%	95%	5%
79	100%	0%	0%	100%	100%	0%

Com base nas informações apresentadas nas Tabelas 16 e 17, é possível observar que as quantidades de referências identificadas pela combinação das técnicas de extração de referências são sempre maiores ou iguais às quantidades identificadas pelas técnicas BibTex e PDF separadamente, conforme Tabela 16. O mesmo vale para os percentuais de acerto apresentados na Tabela 17, o que mostra que a combinação das técnicas, neste estudo, sempre resulta em resultados melhores ou ao menos iguais à aplicação das técnicas isoladamente.

Como o estudo exploratório também visava à comparação entre os algoritmos para cálculo de similaridade de referências de Salton e Allan (1994) e o novo algoritmo proposto, dados foram coletados para tanto.

Antes da aplicação dos algoritmos, foi mapeado o número de referências em duplicidade na pesquisa para as técnicas de extração de referências por BibTex e por PDF. Avaliando todas as referências obtidas por arquivos BibTex, 61 citações apareceram em duplicidade, isso significa dizer que 61 estudos dentre os já conhecidos e as novas referências encontradas foram citados pelo menos uma vez por algum dos estudos já conhecidos. Para os arquivos PDF, foram encontradas 215 duplicidades. Vale destacar que os números são divergentes porque, para alguns estudos, não são exportadas as referências no BibTex, o que é um problema, e no arquivo PDF todas as referências dos

estudos estão presentes, mas nem sempre são reconhecidas pelos algoritmos de processamento de texto.

A Tabela 18 mostra a comparação entre os algoritmos de cálculo de similaridade mencionados ao aplicar as técnicas de extração a partir de arquivos BibTex e de arquivos PDF. A primeira linha da tabela refere-se à quantidade real de referências duplicadas que foram manualmente identificadas pelo autor. As demais linhas referem-se à quantidade de referências duplicadas identificadas pelo algoritmo de Salton e pelo novo algoritmo criado.

Tabela 18. Comparação entre os desempenhos dos algoritmos de comparação de referências

Quantidade de referências	Técnica BibTex	Técnica PDF
Reais existentes no arquivo PDF	61	215
Identificadas corretamente pelo algoritmo de Salton	51	104
Não identificadas como similares pelo algoritmo de Salton	10	111
Identificadas corretamente pelo novo algoritmo	61	215
Não identificadas como similares pelo novo algoritmo	0	0

Com base nos dados apresentados na Tabela 18, é possível observar que o novo algoritmo proposto se mostra superior para a finalidade de reconhecer novos estudos com base nas referências existentes nos arquivos *BibTex* e PDF. Para chegar a essa análise, basta avaliar as linhas de dados 2 e 4 e as linhas 3 e 5, que apresentam respectivamente as duplicações identificadas corretamente pelos algoritmos de Salton e o novo algoritmo e os erros dos algoritmos de Salton e do novo algoritmo. O cabeçalho da tabela não é considerado como linha de dados.

O Algoritmo de Salton perde muito sua efetividade na técnica baseada em PDF (linha 2, coluna 2 comparada à linha 1, coluna 2), isso porque esse método apresenta problemas de reconhecimento de padrões para identificação das referências, podendo gerar referências incompletas ou erradas.

5.3.8.2 Análise Estatística

A Tabela 19 apresenta as médias dos erros e acertos em cada estratégia de extração de referências (por BibTex e por PDF). A primeira e segunda colunas apresentam, respectivamente, o número de acertos e de erros da técnica de extração de referências em PDF. A terceira e quarta colunas apresentam,

respectivamente, o número de acertos e de erros da técnica de extração de referências em BibTex. A quinta e sexta colunas apresentam, respectivamente, o número de acertos e de erros da técnica de extração de referências das duas técnicas combinadas. Ressalta-se que ao somar os acertos e os erros de cada técnica, o valor resultante deve ser 100%.

Tabela 19. Médias de erros e acertos na identificação das referências

PDF		BibTex		PDF + BibTex	
ACERTO	ERRO	ACERTO	ERRO	ACERTO	ERRO
80%	20%	43%	57%	87%	13%

Com relação aos valores apresentados, de forma geral, a técnica PDF teve uma taxa de acerto de 80% para recuperação de referências. A técnica BibTex teve 43% de acerto. As duas técnicas combinadas tiveram 87% de acerto. A média aritmética foi escolhida para dar o mesmo peso em porcentagens de erro na identificação de referências para cada estudo. Isso implica que estudos com muitas referências, das quais poucas foram identificadas não irão influenciar a média mais a média que os artigos com poucas referências.

Outra análise poderia ser feita considerando o total de referências geral da pesquisa e o erro de identificação geral de referências. Não se recomenda esse tipo de análise pela presença de *outliers* no estudo que influenciariam a média.

Com relação aos algoritmos de similaridade, foi calculada a média dos valores de *acerto* dos algoritmos e o percentual de erro envolvido. A técnica BibTex recupera os dados com muito menos ruído (partes de texto reconhecidas incorretamente a mais ou informações que faltaram na referência). Já é conhecido que o algoritmo de Salton é bastante sensível a ruído e, portanto, as técnicas foram avaliadas separadamente.

A Tabela 20 apresenta os valores de acertos e erros para cada um dos algoritmos aplicados (Salton e novo) a cada uma das técnicas (BibTex e PDF).

Tabela 20. Médias de acertos e erros para os algoritmos de similaridade

	SALTON		Novo Algoritmo	
	Técnica BibTex	Técnica PDF	Técnica BibTex	Técnica PDF
Acerto	84%	48%	100%	100%
Erro	16%	52%	0%	0%

O novo algoritmo proposto mostra-se superior para a finalidade de reconhecer novos estudos. Isso pode ser concluído considerando os erros dos algoritmos apresentados na última linha da Tabela 20. O novo algoritmo teve precisão de 100%, o que equivale a dizer que não teve erro algum. O novo algoritmo proposto é passível de falha e falharia nos seguintes cenários: (i) um título de referência contido dentro de outro título de outra referência (citação pouco comum), como por exemplo os títulos “Systematic Review X” e “Systematic Review X: An Update”; ou (ii) caso os dois títulos possuíssem ruídos distintos, como por exemplo os títulos identificados “2017Systematic Review” e “pp4v3.Systematic Review”. Essas situações não aconteceram no conjunto de dados do estudo exploratório realizado, mas não se pode afirmar que não ocorreriam em outras pesquisas.

Com base nos resultados e análises apresentadas, é possível concluir que os algoritmos criados possuem uma boa precisão, porém falham ao não identificar todas as referências existentes e podem falhar ao identificar o número de duplicações e citações internas da pesquisa.

Em resumo, respondendo as questões levantadas na Seção 5.3.4 deste capítulo, temos que, com base na análise conjunta das Tabelas 16, 17 e 19, as técnicas de extração de referências mostraram-se mais efetivas de forma combinada, isto é, considerando tanto o PDF quanto o BibTex, no contexto deste estudo. Além disso, com base na análise conjunta das Tabelas 18 e 20, o novo algoritmo criado apresenta melhores resultados que o algoritmo de Salton para fins de comparação de referências.

5.4 Experimento: a avaliação de critérios de priorização para aplicação de *snowballing backward* em busca com *String*

Wohlin (2014) propõe que o uso de *snowballing* como estratégia primária e única para a execução de revisões sistemáticas é suficiente. Entretanto, existe uma série de problemas relacionados à técnica *snowballing*. O principal desafio é a construção do conjunto inicial de estudos (*seed*). A comunidade normalmente utiliza uma forma alternativa a de Wohlin (2014), aplicando *snowballing* nos resultados

obtidos ao executar uma *string* de busca em máquinas de busca. Kitchenham e Brereton (2013) reportam uma estratégia nesse sentido.

Nesta seção são apresentados os resultados e as análises do experimento realizado, que teve como objetivo investigar a efetividade e eficiência do uso de *snowballing backward* na melhora de resultados obtidos com *string* de busca e a aplicação dos critérios de priorização propostos. As principais etapas do experimento serão relatadas conforme o processo definido por Wohlin *et al.* (2000).

5.4.1 Planejamento e Contexto do Experimento

O experimento foi planejado com dois objetivos principais: verificar se *snowballing* pode ser utilizado para melhorar os resultados de uma *string* de busca e verificar se os critérios de priorização propostos neste trabalho podem ser utilizados para auxiliar o pesquisador a executar a técnica. Um objetivo secundário é avaliar o suporte computacional desenvolvido.

Esse estudo foi planejado e executado utilizando o método GQM (Goal, Question, Metric) (BASILI; 1992). A descrição do estudo é apresentada na Tabela 21.

O experimento foi considerado uma das atividades da disciplina de revisão sistemática do Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Universidade Federal de São Carlos (UFSCar). A disciplina recebeu inscrições de alunos pertencentes a qualquer programa de pós-graduação da universidade, não apenas de computação. Os alunos não receberam uma nota pelo desempenho no resultado, entretanto a participação foi considerada uma atividade da disciplina. Os alunos foram contextualizados sobre a pesquisa e sobre o projeto da ferramenta StArt e que, ao participar do experimento, estariam contribuindo com a pesquisa. Todos os participantes se mostraram dispostos a participar e demonstraram interesse em contribuir e utilizar a técnica na ferramenta StArt em suas pesquisas.

Tabela 21. GQM planejado para o experimento

Analisar	<i>Snowballing</i> aplicada aos resultados de uma busca com <i>String</i> auxiliado pelos critérios de priorização propostos.
Com o propósito de	Avaliar a melhoria dos resultados da busca e a efetividade dos critérios de priorização
Com respeito a	Esforço e precisão dos resultados.
Do ponto de vista dos	Pesquisadores
No contexto de	Estudantes de pós-graduação da Universidade de Federal de São Carlos.

Wohlin *et al.* (2000) definem quatro dimensões que caracterizam o contexto de um experimento. O experimento aqui reportado foi do tipo *off-line* por não ser construído ou baseado no desenvolvimento de software pela indústria; os participantes são estudantes de pós-graduação (mestrado ou doutorado); o problema é do tipo *toy*, representando uma fração da pesquisa de cada participante.

Os resultados obtidos por esse estudo podem auxiliar a refinar e a melhorar as estratégias propostas para *snowballing* bem com a melhorar o suporte provido pela ferramenta StArt.

5.4.2 Seleção das Variáveis

As seguintes variáveis foram coletadas no experimento:

- O número de participantes que encontraram estudos relevantes após aplicar *snowballing backward* em suas pesquisas.
- O número de participantes que encontraram ao menos um estudo relevante ranqueado por cada um dos critérios de priorização existentes individualmente.
- A quantidade de estudos aceitos e ranqueados por cada um dos critérios de priorização utilizados individualmente e em conjunto.
- A quantidade de estudos rejeitados e ranqueados por cada um dos critérios de priorização utilizados individualmente e em conjunto.

5.4.3 Seleção dos Participantes

Para selecionar os participantes, os mesmos foram escolhidos com base na conveniência, de modo a utilizar a disciplina de Revisão Sistemática que estava sendo ministrada aos alunos do PPGCC da UFSCar. *Snowballing* e estudos secundários faziam parte da ementa da disciplina. A maioria dos participantes já teve contato com pesquisa científica e desenvolve seus próprios projetos de pesquisa em conjunto com seus orientadores de diversas áreas de conhecimento.

5.4.4 Projeto Experimental

O experimento foi realizado de forma individual, cada participante realizou o experimento considerando seu próprio tema de pesquisa de pós-graduação. Para uma melhor compreensão, os participantes foram divididos em sete grupos de acordo com a área de pesquisa, sendo elas: ciências humanas e educação, ciência da computação – subárea Engenharia de Software, ciência da computação - subárea inteligência artificial, ciência da computação - outras áreas, engenharia de produção, ciências biológicas e da saúde e, por fim, química e engenharia química.

O estudo foi executado em quatro dias distintos. As atividades duraram quatro horas em cada dia, conforme explicado a seguir:

- Dia 1: Os participantes definiram o tema de pesquisa e construíram o protocolo de suas revisões em comum acordo com seus orientadores de pesquisa.
- Dia 2: Os participantes construíram as *strings* de busca e refinaram a busca para retornar resultados melhores.
- Dia 3: Os participantes executam *snowballing* na StArt para uma amostra dos resultados retornados pela *string* de busca e foram instruídos a realizar a priorização das referências e a realizar a avaliação das referências encontradas na amostra, além de responderem um breve questionário.
- Dia 4: Os participantes classificaram uma amostra dos estudos encontrados com a *string* de busca. As amostras utilizadas foram escolhidas utilizando os quadrantes propostos por Octaviano et al. (2015) e Octaviano et al. (2016), sendo selecionados os estudos pertencentes ao primeiro e quarto quadrantes por conterem maior número de estudos aceitos e rejeitados, respectivamente. Assim, a amostra selecionada foi heterogênea. Todas as referências encontradas para os estudos da

amostra da pesquisa foram classificadas. Os participantes entregaram os dados classificados no formato utilizado pela ferramenta (.start) para que os dados pudessem ser avaliados futuramente.

Os dias 1 e 2 envolveram treinamento. Nos dias 3 e 4 foram coletados dados. Nem todos os participantes participaram de todos os dias do estudo. A Tabela 22 apresenta a distribuição de participantes por grupo (coluna 1), a sigla de cada grupo (coluna 2) bem como o número de participantes que participaram dos dias 3 e 4 (colunas 3 e 4 respectivamente). As linhas representam os grupos definidos para os participantes.

Tabela 22. Participantes por área de conhecimento

GRUPO	SIGLA	Nº de Participantes (dia 3)	Nº de Participantes (dia 4)
CIÊNCIAS HUMANAS E EDUCAÇÃO	HE	4	1
CIENCIA DA COMPUTAÇÃO	CES	8	
ENGENHARIA DE SOFTWARE			7
CIENCIA DA COMPUTAÇÃO	CIA	3	
INTELIGÊNCIA ARTIFICIAL			2
CIENCIAS DA COMPUTAÇÃO	CC	5	
OUTRAS ÁREAS			3
ENGENHARIA DE PRODUÇÃO	EP	6	5
CIENCIAS BIOLOGICAS E DA SAÚDE	CBS	5	3
QUIMICA E ENGENHARIA QUÍMICA	EQ	4	3

5.4.5 Instrumentação

Para a realização do experimento foi utilizada uma versão construída para o experimento da ferramenta StArt, a qual possuía recursos para aplicação de *snowballing* e dos critérios de priorização propostos. Além da ferramenta, os participantes utilizaram a SCOPUS para realizar pesquisas bibliográficas.

5.4.6 Questões a serem respondidas

As seguintes questões foram definidas:

- Critérios de priorização podem ser utilizados para a classificação automática de todo o conjunto de estudos encontrados?
- Critérios de priorização podem ser utilizados para a redução do volume de estudos a serem processados pelo pesquisador?
- Em qual ordem pode-se aplicar os critérios de priorização?

-
- Snowballing backward pode melhorar os resultados de uma busca com *strings*?
 - É possível encontrar estudos relevantes aplicando *snowballing* em estudos rejeitados?
 - O suporte computacional desenvolvido mostra indícios de reduzir o esforço ao aplicar *snowballing*?

5.4.7 Ameaças à Validade

A seguir são apresentados alguns itens que poderiam representar ameaças à validade do experimento realizado.

- **Validade Interna:**
 1. Desempenho e vontade do participante pelo fato do experimento ter sido realizado durante uma disciplina de pós-graduação: como as notas foram atribuídas pela participação do estudante, considera-se que o risco associado ao desempenho tenha sido minimizado.
 2. Erro no registro dos dados coletados: esse risco não pode ser minimizado pois cada participante foi responsável pela execução das buscas e importação na ferramenta. Os participantes deveriam salvar os dados ao final das etapas de execução do estudo.
- **Validade Externa**
 1. Os resultados não podem ser generalizados para outros perfis de pesquisas. Essa ameaça não pode ser minimizada, pois não houve a possibilidade de aplicar o experimento com outros tipos de pesquisadores e com diferentes níveis de experiência.
 2. Os grupos foram determinados em acordo com a área de pesquisa, e a quantidade de estudos dependeu dos resultados encontrados na busca. É possível que ao utilizar outras bases de dados ou mudar as buscas, o resultado seja diferente.
 3. A execução do estudo em momentos futuros implica em resultados diferentes na execução da busca executadas nas máquinas de busca e, conseqüentemente, a amostra de estudos avaliados seria diferente.

- **Validade de Conclusão**

1. Como apenas parte do conjunto de estudos da pesquisa foi utilizado, é possível que os resultados fossem diferentes com o uso de uma amostra diferente de estudos.
2. Uma possível ameaça são os algoritmos de processamento de texto utilizados na ferramenta StArt, os quais podem conter erros. Um estudo foi conduzido para mensurar os erros da ferramenta, entretanto, para medir o valor exato do erro existente, seria necessário executar todo o processo manualmente.

5.4.8 Método

As seguintes tarefas foram realizadas:

- Tarefa 1: preencher um questionário de caracterização, o qual incluía informações sobre detalhes pessoais e técnicos (dia 1).
- Tarefa 2: Todos os participantes receberam um treinamento sobre RS e sobre a ferramenta StArt. Esse treinamento visou ao nivelamento técnico dos participantes (dia 1).
- Tarefa 3: Todos os participantes construíram um protocolo completo para uma revisão sistemática em seu tema de pesquisa (dia 2).
- Tarefa 4: Todos os participantes receberam treinamento sobre buscas com *string* e sobre a máquina de busca SCOPUS. Esse treinamento foi necessário, pois a SCOPUS nem sempre é utilizada com frequência em pesquisas que não são da área da computação (dia 3).
- Tarefa 5: Os participantes construíram as *strings* de busca e realizaram a busca na Scopus. As buscas foram importadas na ferramenta StArt (dia 3).
- Tarefa 6: Os participantes receberam um treinamento prévio sobre *snowballing*, realizaram *snowballing* manual em uma amostra reduzida de 5 estudos do seu tema de pesquisa. Após realizado *snowballing* manual, foram apresentados aos participantes recursos para *snowballing* desenvolvidos na ferramenta StArt (dia 3).
- Tarefa 7: os participantes foram instruídos a localizar um único estudo relevante, caso possível, ranqueado por score, número de citações,

frequência de autores ou frequência de referências disponível na Start. Após realizada a atividade os participantes responderam um breve questionário (dia 3).

- Tarefa 8: Os participantes classificaram uma amostra dos estudos, ranqueada pelos quadrantes da estratégia SCAS para seleção semiautomática de estudos primários (Octaviano et al., 2015). Depois disso, os participantes classificaram todas as referências identificadas por *snowballing* pela StArt da amostra classificada inicialmente (dia 4).
- Tarefa 9: Os dados dos participantes foram encapsulados em um arquivo .start para análise. Os dados foram analisados e os resultados são descritos nessa seção (realizada posteriormente ao experimento).

5.4.9 Coleta de Dados

Os dados foram coletados por meio de questionários preenchidos por cada participante do experimento. Além dos questionários, cada participante entregou o arquivo .start com os dados coletados para análise futura.

5.4.10 Análise e Interpretação

As análises foram feitas com suporte da ferramenta Microsoft Excel e os resultados são apresentados a seguir.

5.4.10.1 Análise Descritiva

A Tabela 23 apresenta os dados obtidos com os participantes que estiveram presentes até o terceiro dia do experimento. As colunas 2, 3, 4, 5, 6, 7 e 8 apresentam a quantidade de participantes por grupos, conforme definido na Tabela 22. Na coluna 9 são apresentados os totais de participantes de todos os grupos somados. A primeira linha da tabela mostra o total de participantes por grupo. A segunda e terceira linhas mostram, respectivamente quantos participantes encontraram ou não estudos relevantes para seus tópicos de pesquisa. A quarta, quinta e sexta linhas mostram o número de participantes que encontraram ao menos um estudo relevante com score maior que zero, número de citações maior que zero, score de autores maior que zero e score de conferências maior que zero,

respectivamente. Elas são relacionadas aos critérios de priorização propostos neste trabalho. Por fim, a última linha mostra o total de participantes que consideraram o suporte computacional benéfico na condução de seus estudos secundários.

Tabela 23. Relação de resultados por grupo

Descrição	Grupo							Total
	HE	CES	CIA	CC	EP	CBS	EQ	
Nº Total de participantes	4	9	3	5	6	5	4	35
Nº de participantes que encontraram estudos relevantes	4	9	3	5	6	5	4	35
Nº de participantes que não encontraram estudos relevantes	0	0	0	0	0	0	0	0
Nº de participantes que encontraram ao menos um estudo relevante com Score maior que 0	4	9	3	5	6	5	4	35
Nº de participantes que encontraram ao menos um estudo relevante com Número de citações maior que 0	4	8	3	5	5	4	3	31
Nº Número de participantes que encontraram ao menos um estudo relevante com ranking de autores maior que a média	4	9	2	5	5	4	1	30
Nº de participantes que encontraram ao menos um estudo relevante com ranking de conferências maior que 0	4	7	3	4	5	3	1	27
Nº de participantes que apontam benefício ao utilizar suporte computacional	4	9	3	5	5	4	4	35

Com relação aos resultados apresentados na Tabela 23, na segunda linha de dados (não considerando o cabeçalho como linha de dados) é possível verificar que todos os participantes encontraram novos estudos relevantes utilizando *snowballing backward* a partir dos resultados da busca com *strings*. Mesmo com a técnica *snowballing* sendo executada apenas em uma amostra dos estudos, foram encontrados novos estudos relevantes além dos estudos anteriormente identificados por meio de *strings* de busca.

A partir das análises das linhas de dados 4, 5, 6 e 7, é possível observar que os critérios de priorização se mostram efetivos. O número de participantes que encontraram ao menos um estudo relevante com *score* maior que 0 foi 35 (100%). O número de participantes que encontraram ao menos um estudo relevante com número de citações maior que 0 foi 31 (89%). O número de participantes que encontraram ao menos um estudo relevante com *ranking* de autores maior que a média foi 30 (86%). O Número de participantes que encontraram ao menos um estudo relevante com *ranking* de conferências maior que 0 foi 27 (77%). A última linha da Tabela 23 mostra o número de participantes que considera benéfico o uso de apoio computacional para aplicação de *snowballing*. Todos os participantes (100%) consideraram benéfico.

A Tabela 24 apresenta os dados quantitativos de cada participante dos grupos ao aplicar inicialmente a técnica busca por *string* e, posteriormente, o *snowballing* no experimento. Esses dados foram coletados no quarto dia. A primeira coluna apresenta a identificação do participante, a segunda coluna apresenta o grupo definido de acordo com o tema de pesquisa. A terceira coluna apresenta o número de estudos identificados pela *string* de busca do participante. A quarta coluna apresenta a amostra classificada pelo participante dos estudos encontrados pela *string* de busca. A quinta coluna apresenta o número de novos estudos (relevantes ou não) identificados pela aplicação de *snowballing* na amostra da quarta coluna. A sexta coluna apresenta o número de duplicações identificadas automaticamente pela ferramenta na aplicação de *snowballing* dos estudos da amostra da quarta coluna. A sétima coluna apresenta o número de estudos não relevantes encontrados por *snowballing* (apresentados na quinta coluna). A oitava coluna apresenta o número de estudos relevantes encontrados por *snowballing* (apresentados na quinta coluna). Com base na tabela, podemos avaliar o número de novos estudos relevantes

encontrados por cada participante. Também podemos observar a grande quantidade de duplicações (identificada automaticamente pela ferramenta).

Tabela 24. Dados das pesquisas dos participantes

Participante	Grupo	String de busca		Snowballing			
		Identificados	Classificados	Identificados	Duplicados	Rejeitados	ACEITOS
1	CBS	139	50	3812	1689	1254	108
2	EP	60	51	858	458	747	26
3	CES	314	50	2251	4104	326	41
4	EQ	20	20	327	273	305	22
5	CES	179	53	3104	797	990	10
6	CBS	11	11	182	67	169	13
7	HE	160	50	191	123	95	9
8	CC	292	50	3940	2074	654	36
9	CES	266	100	59	5	13	11
10	EP	633	50	12631	8543	1243	34
11	CES	23	23	200	168	163	37
12	EQ	79	50	1674	345	686	116
13	EP	329	50	5784	3227	737	56
14	CC	139	50	2218	1063	964	60
15	CES	164	50	1774	1557	653	17
16	CBS	75	75	2270	694	2238	32
17	CIA	110	50	683	117	351	30
18	CES	166	50	3466	2113	1318	123
19	CES	112	50	1735	738	713	14
20	EP	400	48	8597	5716	1001	52
21	EQ	31	722	172	894	55	576
22	EP	51	3115	1118	4233	65	909
23	CIA	50	74	6	80	9	36
24	CC	50	3117	2822	5939	39	506

Nota-se o crescimento exponencial de *snowballing* observando-se os valores da terceira e quinta colunas. O número de estudos encontrados para serem processados pelo pesquisador cresce de forma acentuada em apenas uma iteração de *snowballing backward*. Com a análise conjunta das duas últimas colunas, observa-se a grande quantidade de estudos que são encontrados por *snowballing* e rejeitados, em detrimento de poucos estudos aceitos e que serão relevantes para a pesquisa, sendo esses contabilizados na última coluna.

A utilização de apoio computacional mostrou-se positiva, como pode ser observado na coluna 6 que indica o número de duplicações identificadas

automaticamente na ferramenta ao aplicar *snowballing*. Isso resulta na redução de esforço do pesquisador para execução dessa tarefa.

Os números referentes às duplicações de estudos identificadas por participante são mostrados na Tabela 25. As colunas 2 e 3 indicam, respectivamente, o número de estudos não duplicados (únicos) e o número de duplicados identificados. A coluna 4 mostra o percentual de duplicações identificado em relação ao total de estudos (soma das colunas 2 e 3). É possível calcular que, em média, o esforço foi reduzido em 32% ao detectar de forma automática as duplicações de estudos depois de aplicar *snowballing*, como mostrado na última linha da tabela.

Tabela 25. Porcentagem de estudos duplicados encontrados

Participante	Total ÚNICO	Duplicações	Porcentagem de Duplicações
1	3812	1689	31%
2	858	458	35%
3	2251	4104	65%
4	327	273	46%
5	3104	797	20%
6	182	67	27%
7	191	123	39%
8	3940	2074	34%
9	59	5	8%
10	12631	8543	40%
11	200	168	46%
12	1674	345	17%
13	5784	3227	36%
14	2218	1063	32%
15	1774	1557	47%
16	2270	694	23%
17	683	117	15%
18	3466	2113	38%
19	1735	738	30%
20	8597	5716	40%
21	722	172	19%
22	3115	1118	26%
23	74	6	8%
24	3117	2822	48%
MEDIA:			32%

A Tabela 26 apresenta os totais gerais de referências encontradas, os totais de duplicações e a porcentagem geral de duplicações encontradas. A porcentagem geral de duplicações é de 38%. Isso pode indicar que existem pesquisas com muitos

estudos nas quais as porcentagens de duplicações são maiores e pesquisas com poucos estudos, com porcentagem de duplicação menores. Isso influencia o valor da média apresentada. As porcentagens de duplicações da média foram 32%, indicando uma maior frequência de pesquisas com menor número de duplicações.

Tabela 26. Total geral de referências e duplicações do estudo

Total de referências encontradas (Geral)	Total de Duplicações encontradas (Geral)	Porcentagem de Duplicações (Geral)
62784	37989	38%

Outro benefício do apoio computacional é em relação à coleta das referências, que é feita de modo automático pela ferramenta. Assim, o pesquisador não tem a necessidade de localizá-las manualmente, o que demandaria esforço maior. Concluída a parte prática do experimento, os pesquisadores responderam a seguinte pergunta: o suporte implementado na ferramenta StArt auxilia na execução do *snowballing*? Havia quatro respostas possíveis: “Ajuda muito pois reduz muito tempo e esforço gastos”, “Ajuda bem pois reduz o tempo e o esforço gastos”, “Ajuda pouco” e “Não ajuda”. Houve 15 respostas referentes à primeira opção, 20 respostas referentes à segunda opção e nenhuma resposta referente às terceira e quarta opções, como apresentado na Figura 15. Os números são muito favoráveis no que diz respeito ao suporte existente na ferramenta para aplicação de *snowballing*.

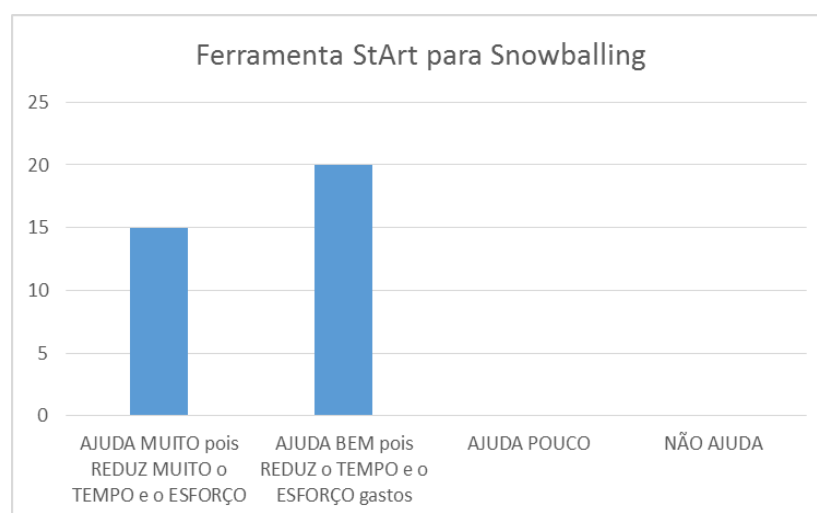


Figura 15. Avaliação dos participantes sobre os recursos de snowballing na StArt

Ressalta-se que os participantes realizaram *snowballing* manualmente apenas para apenas um conjunto restrito de cinco estudos durante o treinamento.

Os participantes conduziram o restante do experimento utilizando apenas apoio computacional. Para uma análise mais aprofundada sobre a avaliação de apoio computacional seria necessário que os participantes tivessem realizado o mesmo esforço para cada uma das técnicas (manual e com o apoio computacional). E assim existiriam resultados possivelmente mais confiáveis.

Com relação aos estudos nos quais a técnica *snowballing* foi aplicada, foi realizada uma comparação entre a quantidade de estudos encontrados por *snowballing* a partir de estudos aceitos e por *snowballing* a partir de estudos rejeitados. Os dados são apresentados na Tabela 27. Na primeira coluna é apresentada a identificação do participante. Na segunda coluna é apresentada a amostra classificada pelo participante dos resultados da busca com strings. Na terceira coluna é apresentado o número de estudos aceitos na amostra pelo participante. Na quarta coluna é apresentado o número de estudos rejeitados pelo participante. Na quinta coluna é apresentado o número de estudos relevantes após aplicada uma iteração de *snowballing backward* em todos os estudos classificados da amostra. Na sexta, sétima, oitava e nona colunas, são apresentados respectivamente, o número de estudos relevantes encontrados por *snowballing backward* dos estudos classificados como aceitos na amostra, a porcentagem relativa desses estudos na quantidade total de estudos relevantes encontrados, o número de estudos relevantes encontrados por *snowballing backward* dos estudos classificados como rejeitados na amostra, e a porcentagem relativa desses estudos na quantidade total de estudos relevantes encontrados.

Por exemplo, para o participante (1) na primeira linha de dados da tabela, o participante realizou a classificação de 50 estudos provenientes de uma busca com strings. Desses estudos, o participante aceitou 10 e rejeitou 40. Ao aplicar *snowballing backward* nos 50 estudos classificados, foram encontrados um total de 108 novos estudos relevantes ainda não identificados pela busca com strings. Desse total, 66 estudos foram obtidos por *snowballing backward* a partir de estudos aceitos (61% do total) e 42 estudos foram obtidos por *snowballing* a partir de estudos rejeitados (39% do total).

Tabela 27. Relação de estudos encontrados por *snowballing* para estudos aceitos ou rejeitados

Participante	Busca com <i>string</i>			Snowballing				
	Amostra	Aceitos	Rejeitados	Total de estudos relevantes identificados por <i>snowballing</i>	Gerados por Snowballing de estudos Aceitos		Gerado por Snowballing de estudos rejeitados	
					Nº	(%)	Nº	(%)
1	50	10	40	108	66	61%	42	39%
2	51	5	46	26	24	92%	2	8%
3	50	33	17	41	37	90%	4	10%
4	20	3	17	22	8	36%	14	64%
5	53	11	42	10	9	90%	1	10%
6	11	6	5	13	13	100%	0	0%
7	50	14	36	9	0	0%	9	100%
8	50	21	29	36	21	58%	15	42%
9	100	25	75	11	11	100%	0	0%
10	50	14	36	34	28	82%	6	18%
11	23	17	6	37	35	95%	2	5%
12	50	14	36	116	26	22%	90	78%
13	50	23	27	56	43	77%	13	23%
14	50	36	14	60	56	93%	4	7%
15	50	10	40	17	10	59%	7	41%
16	75	9	41	32	5	16%	27	84%
17	50	5	45	30	0	0%	30	100%
18	50	23	27	123	26	21%	97	79%
19	50	12	38	14	7	50%	7	50%
20	48	10	38	52	30	58%	22	42%
21	722	17	14	55	38	69%	17	31%
22	3115	25	26	65	33	51%	32	49%
23	74	22	28	9	9	100%	0	0%
24	3117	15	35	39	33	85%	6	15%

Média: 63% Média: 37%

Com base nos valores apresentados na Tabela 27, *snowballing backward* a partir de estudos aceitos é mais efetivo para encontrar novos estudos relevantes (em média representou 63% dos resultados), entretanto *snowballing backward* a partir de estudos rejeitados também localizou estudos relevantes (em média, 37% dos resultados).

Ressalta-se que, na prática, para aplicação de *snowballing* em todos os estudos, incluindo os rejeitados, o esforço cresce enormemente e a definição de critérios de priorização de estudos pode auxiliar e muito na redução do esforço.

5.4.10.2 Análise Estatística

Os critérios de priorização propostos também foram avaliados com base no resultado entregue pelos grupos. Foi aplicado o teste do coeficiente Kappa proposto por Carletta (1996), na tentativa de verificar o nível de concordância entre a revisão manual e uma possível revisão automática utilizando apenas os critérios propostos.

O cálculo do kappa é realizado pela equação: $k = (Pr(a) - Pr(e)) / (1 - Pr(e))$, na qual, $Pr(a)$ é o acordo relativo entre os critérios e participantes e $Pr(e)$ é a probabilidade de acordo entre as classificações, calculada utilizando o comportamento dos dados observados.

Para avaliação dos resultados, pode-se utilizar a tabela proposta por Landis e Koch (1977) e apresentada na Tabela 28.

Tabela 28. Tabela de interpretação do Kappa

Valor Kappa	Interpretação
<0	<i>No agreement</i>
0.00-0.19	<i>Poor agreement</i>
0.20-0.39	<i>Fair agreement</i>
0.40-0.59	<i>Moderate agreement</i>
0.60-0.79	<i>Substantial agreement</i>
0.80-1.00	<i>Almost perfect agreement</i>

A Tabela 29 apresenta os resultados do aplicar o coeficiente Kappa para medir o nível de concordância entre a classificação manual feita pelos participantes e a classificação obtida por meio dos critérios de priorização propostos para *snowballing*. A fim de avaliar os critérios mais efetivos, os quatro critérios de priorização - *score* (SC), número de citações (CIT), autores (AU) e local de publicação (LO) - foram analisados individualmente, como apresentado nas colunas 2 a 5, e também combinados entre si, conforme apresentado nas colunas 6 a 16, gerando novos critérios. Ao todo foram medidos 15 critérios (4 individuais e 11 combinações), que são identificados por: SC, CIT, AU, LO, OU (SC + CIT + LO + OU), SC+CIT, SC+AU, SC+LO, CIT+AU, CIT+LO, AU+LO, SC+CIT+AU, SC+CIT+LO, SC+AU+LO, e CIT+AU+LO. As cores representam o nível de concordância entre a avaliação manual do pesquisador e a avaliação feita por meio dos critérios propostos: *no agreement* (rosa), *poor agreement* (laranja), *fair agreement* (verde), *moderate agreement* (azul), e *substantial agreement* (cinza).

Tabela 29. Aplicação do coeficiente Kappa para os critérios propostos

P	SC	CIT	AU	LO	OU	SC + CIT	SC+AU	SC+LO	CIT + AU	CIT+LO	AU+LO	SC+CIT+AU	SC+CIT+LO	SC+AU+LO	CIT+AU+LO
1	0,31	0,12	0,13	0,04	0,13	0,25	0,25	0,18	0,21	0,13	0,09	0,10	0,16	0,18	0,14
2	0,18	-0,03	-0,03	0,01	0,02	0,14	0,14	0,03	0,09	-0,03	0,00	-0,02	0,03	0,08	0,03
3	0,64	0,12	0,13	0,03	0,19	0,34	0,34	0,28	0,49	0,13	0,12	0,15	0,20	0,29	0,25
4	0,12	-0,03	-0,03	0,04	0,04	0,11	0,11	0,05	0,07	-0,04	0,02	-0,01	0,05	0,06	0,04
5	0,09	-0,02	0,02	0,00	0,02	0,06	0,06	0,02	0,04	0,02	-0,01	0,01	0,02	0,03	0,02
6	0,25	0,27	0,29	0,04	0,13	0,23	0,23	0,18	0,20	0,28	0,15	0,22	0,16	0,18	0,15
7	0,32	0,16	-0,03	-0,12	0,04	0,30	0,30	0,10	0,09	0,02	-0,04	-0,07	0,10	0,08	0,05
8	0,08	0,00	-0,05	-0,05	-0,02	0,03	0,03	-0,02	0,00	-0,03	-0,02	-0,06	-0,02	0,00	-0,03
9	0,59	0,00	0,15	-0,08	0,36	0,59	0,59	0,43	0,51	0,15	-0,08	0,07	0,43	0,51	0,36
10	0,07	0,09	0,06	0,01	0,04	0,07	0,07	0,05	0,04	0,05	0,06	0,05	0,05	0,06	0,04
11	0,38	0,08	0,13	0,11	0,17	0,25	0,25	0,24	0,34	0,11	0,16	0,13	0,18	0,23	0,23
12	0,40	-0,02	-0,04	-0,01	0,14	0,36	0,36	0,17	0,28	-0,04	-0,02	-0,01	0,16	0,26	0,14
13	0,31	0,14	-0,02	-0,01	0,08	0,27	0,27	0,11	0,17	0,01	0,05	-0,02	0,11	0,16	0,09
14	0,07	0,08	0,06	-0,02	0,03	0,07	0,07	0,04	0,06	0,05	0,04	0,03	0,04	0,06	0,03
15	0,08	-0,04	0,02	-0,02	0,04	0,05	0,05	0,04	0,06	0,01	-0,03	0,01	0,04	0,04	0,04
16	0,02	0,06	0,02	0,01	0,01	0,02	0,02	0,01	0,01	0,02	0,04	0,02	0,01	0,01	0,01
17	0,19	0,21	0,05	0,00	0,09	0,18	0,18	0,10	0,16	0,08	0,14	0,05	0,10	0,15	0,09
18	0,11	0,03	0,04	0,03	0,05	0,08	0,08	0,07	0,09	0,05	0,03	0,04	0,06	0,06	0,07
19	0,03	-0,03	0,02	-0,01	0,01	0,02	0,02	0,01	0,01	0,01	-0,02	0,01	0,01	0,01	0,01
20	0,18	0,02	0,08	0,04	0,07	0,16	0,16	0,10	0,13	0,07	0,03	0,06	0,09	0,12	0,08
21	0,07	0,16	0,02	0,03	0,05	0,10	0,10	0,03	0,06	0,05	0,11	0,03	0,04	0,09	0,04
22	0,38	0,21	0,02	-0,02	0,10	0,28	0,28	0,16	0,20	0,08	0,10	0,01	0,14	0,17	0,11
23	0,54	0,12	0,05	0,06	0,29	0,49	0,49	0,42	0,31	0,12	0,14	0,12	0,39	0,28	0,32
24	0,16	0,02	-0,02	-0,02	0,07	0,10	0,10	0,11	0,12	0,01	0,01	-0,03	0,08	0,08	0,09

Com relação aos resultados, é possível observar que a avaliação manual divergiu na maioria dos casos quando comparada aos critérios propostos, indicando que os critérios aplicados mostraram nenhuma concordância (rosa) ou baixa concordância (laranja) em relação à revisão manual. Os resultados negativos do teste podem ser explicados pelo fato de os critérios terem um número elevado de falsos negativos. Isso apresenta indícios de que esses critérios não devem ser tratados para classificação automática de todo o conjunto de estudos.

Considerando os estudos não ranqueados pelos critérios propostos, o número de referências da revisão a ser avaliado torna-se bastante elevado e, pelo fato de não ser ranqueado em nenhuma das classificações de priorização, a tendência é que a grande maioria dessas referências não sejam relevantes. Assim, o número de estudos que podem ser filtrados na revisão por meio dos critérios de priorização é grande. Aplicar os critérios de priorização pode reduzir muito o volume de

referências a serem processadas. Vale ressaltar que, ao aplicar o critério para reduzir o volume de referências, existe o risco de descartar algumas que sejam relevantes.

Em uma segunda análise, os critérios foram avaliados para rejeitar estudos não relevantes. Nessa nova estratégia, propõe-se que estudos que não atendam a nenhum dos critérios propostos sejam rejeitados e, assim, o pesquisador não terá de revisar as referências que não atenderem a nenhum critério de priorização. Os resultados foram avaliados por meio de *precision* e *recall*.

No contexto deste experimento, define-se as métricas *precision* e *recall* da seguinte forma: Se o valor de *precision* for 100%, então significa que todos as referências não ranqueadas por nenhum dos critérios de priorização são realmente rejeitadas pelo participante, embora ainda possam existir referências que o participante rejeitaria, mas que foram marcadas com ao menos um dos critérios de priorização. Se o valor de *recall* for 100%, então todos as referências rejeitadas pelo participante também não foram ranqueadas por nenhum dos critérios de priorização, embora possa haver falso-positivos (estudos que seriam aceitos por uma avaliação do participante, mas foram rejeitados por não atenderem a nenhum critério de priorização).

As métricas *precision* e *recall* foram definidas como segue:

$$precision = \frac{|R \cap r|}{|r|}$$

onde:

$$recall = \frac{|R \cap r|}{|R|}$$

$R = \{R \mid R \in \text{estudos não ranqueados por nenhum dos critérios de priorização}\}$

$r = \{t \mid t \in \text{estudos rejeitados manualmente pelo pesquisador}\}$

Para computar as métricas *precision* e *recall*, os participantes realizaram a pré-seleção de todas as referências identificadas na ferramenta StArt manualmente. Com base nos dados de aceitação e rejeição realizadas, os critérios foram aplicados nos estudos e as classificações comparadas. Ressalta-se que os critérios foram aplicados da seguinte forma: caso o estudo atenda a ao menos um dos critérios, o mesmo não recebe nenhuma classificação. Caso o estudo não atenda a nenhum dos critérios, o mesmo é rejeitado.

Como exemplo para a interpretação da Tabela 29 podemos considerar o participante (1). Os dados para cada critério desse participante são apresentados na primeira linha de dados. Para o Critério SC (score), A precisão foi de 96.72% e o recall de 84.53%. O valor de 96.72% de precisão pode ser entendido da seguinte forma: caso o critério de *score* seja utilizado para rejeição automática de todas as referências que não sejam ranqueadas por ele, 96.72% das referências que serão rejeitadas, serão rejeitadas corretamente e 3.28% representam as referências que serão rejeitadas de forma equivocada, ou seja, referências que seriam aceitas na revisão manual do participante. Para o valor de *recall*, de 84.53%, o mesmo pode ser entendido como: 84.53% de todos os estudos que seriam rejeitados em uma revisão manual, de fato foram rejeitados. Para o restante, os 15.47%, os mesmos permanecerão sem classificação.

A mesma interpretação vale para todas as colunas dos critérios. Destaca-se que esses critérios estão sendo avaliados referentes apenas à rejeição e não em relação à aceitação. O erro referente à rejeição de estudos aceitos será retomado na Tabela 33, seguido por devidas análises.

Na Tabela 30 são apresentados os resultados individuais de *precision* e *recall* dos participantes. A primeira coluna apresenta a identificação do participante que esteve presente no quarto dia do estudo. A segunda coluna apresenta o método avaliativo aplicado (*precision* ou *recall*), A terceira coluna apresenta os resultados para o método avaliativo para o critério *score*. A quarta coluna apresenta os resultados para o método avaliativo para o critério de citações internas no estudo secundário. A quinta coluna apresenta os resultados para o método avaliativo para o critério ranking de autores. A sexta coluna apresenta os resultados para o método avaliativo para o critério ranking de conferências. A sétima coluna apresenta a combinação dos quatro critérios de priorização de referências anteriores. As colunas 8 a 13 apresentam os critérios de priorização combinados dois a dois, e as colunas 14 a 17 apresentam os critérios combinado três a três.

Tabela 30. *Precision* e *recall* para os critérios aplicados na rejeição de estudos

P	Método	SC	CIT	AU	LO	OU	SC + CIT	SC+ AU	SC+ LO	CIT + AU	CIT+ LO	AU+ LO	SC+CIT+ AU	SC+CIT+ LO	SC+AU+ LO	CIT+AU+ LO
1	<i>PRECISION</i>	96,72 %	93,29 %	94,57 %	92,45 %	97,91 %	97,20 %	97,48 %	96,80 %	95,34 %	93,67 %	94,73 %	97,89%	97,29%	97,55%	95,44%
	<i>RECALL</i>	84,53 %	89,87 %	77,83 %	87,88 %	55,90 %	77,43 %	67,78 %	74,80 %	71,77 %	79,03 %	68,74 %	62,84%	68,58%	60,21%	63,48%
2	<i>PRECISION</i>	97,63 %	96,55 %	96,17 %	96,69 %	97,51 %	97,57 %	97,38 %	97,89 %	96,06 %	96,59 %	96,11 %	97,31%	97,83%	97,58%	95,99%
	<i>RECALL</i>	84,53 %	89,87 %	77,83 %	87,88 %	55,90 %	77,43 %	67,78 %	74,80 %	71,77 %	79,03 %	68,74 %	62,84%	68,58%	60,21%	63,48%

	RECALL	93,71 %	97,46 %	73,90 %	85,94 %	57,70 %	91,43 %	69,61 %	80,59 %	71,89 %	83,40 %	62,92 %	67,87%	78,31%	59,44%	60,91%
3	PRECION	96,85 %	90,94 %	91,58 %	89,09 %	97,88 %	97,28 %	96,36 %	97,25 %	93,24 %	91,60 %	92,52 %	97,06%	97,86%	96,96%	94,63%
	RECALL	94,17 %	80,06 %	76,69 %	92,64 %	56,75 %	76,69 %	73,01 %	86,81 %	63,50 %	73,62 %	72,09 %	60,74%	70,25%	68,40%	59,51%
	PRECION	95,29 %	93,15 %	92,68 %	93,66 %	95,06 %	95,22 %	95,26 %	95,05 %	92,53 %	93,55 %	93,06 %	95,16%	94,95%	95,18%	92,89%
4	RECALL	79,67 %	98,03 %	74,75 %	87,21 %	50,49 %	78,36 %	59,34 %	69,18 %	73,11 %	85,57 %	65,90 %	58,03%	67,87%	51,80%	64,26%
	PRECION	99,77 %	98,93 %	99,56 %	98,98 %	99,80 %	99,75 %	99,83 %	99,73 %	99,53 %	98,91 %	99,50 %	99,82%	99,72%	99,81%	99,47%
	RECALL	86,86 %	93,33 %	68,45 %	87,97 %	49,85 %	81,40 %	59,96 %	75,53 %	64,81 %	82,51 %	60,16 %	57,13%	71,18%	52,17%	57,13%
6	PRECION	100,00 %	94,71 %	97,26 %	93,21 %	100,00 %	100,00 %	100,00 %	100,00 %	97,84 %	94,74 %	97,69 %	100,00%	100,00 %	100,00 %	98,39%
	RECALL	69,82 %	95,27 %	84,02 %	89,35 %	52,07 %	67,46 %	59,76 %	63,31 %	80,47 %	85,21 %	75,15 %	57,40%	60,95%	54,44%	72,19%
	PRECION	93,81 %	92,16 %	90,67 %	89,89 %	93,10 %	93,75 %	94,20 %	92,77 %	91,78 %	90,80 %	89,23 %	94,12%	92,68%	93,22%	90,48%
7	RECALL	95,79 %	98,95 %	71,58 %	84,21 %	56,84 %	94,74 %	68,42 %	81,05 %	70,53 %	83,16 %	61,05 %	67,37%	80,00%	57,89%	60,00%
	PRECION	95,24 %	94,80 %	93,74 %	94,34 %	93,92 %	95,21 %	94,26 %	94,84 %	93,90 %	94,41 %	93,26 %	94,32%	94,85%	93,84%	93,48%
	RECALL	94,80 %	86,39 %	73,24 %	89,14 %	56,73 %	82,11 %	70,34 %	84,25 %	65,90 %	77,52 %	65,60 %	63,46%	73,24%	62,84%	59,17%
9	PRECION	90,00 %	54,17 %	60,00 %	52,17 %	85,71 %	90,00 %	87,50 %	88,89 %	60,00 %	52,17 %	57,14 %	87,50%	88,89%	85,71%	57,14%
	RECALL	69,23 %	100,00 %	69,23 %	92,31 %	46,15 %	69,23 %	53,85 %	61,54 %	69,23 %	92,31 %	61,54 %	53,85%	61,54%	46,15%	61,54%
	PRECION	98,62 %	98,70 %	98,60 %	97,40 %	99,82 %	99,62 %	99,43 %	98,48 %	99,10 %	98,66 %	98,69 %	99,84%	99,58%	99,37%	99,13%
10	RECALL	74,58 %	79,40 %	73,61 %	90,59 %	44,73 %	63,31 %	56,07 %	67,58 %	61,87 %	71,20 %	66,53 %	50,04%	56,88%	50,68%	55,19%
	PRECION	90,07 %	82,84 %	84,31 %	82,80 %	90,63 %	89,31 %	90,60 %	90,21 %	84,85 %	84,71 %	84,72 %	90,20%	89,52%	90,99%	85,48%
	RECALL	83,44 %	85,89 %	79,14 %	94,48 %	53,37 %	71,78 %	65,03 %	79,14 %	68,71 %	81,60 %	74,85 %	56,44%	68,10%	61,96%	65,03%
12	PRECION	92,79 %	85,33 %	84,69 %	85,40 %	92,31 %	92,63 %	91,75 %	93,19 %	84,63 %	85,25 %	85,08 %	91,63%	93,05%	92,44%	85,01%
	RECALL	86,30 %	97,52 %	72,59 %	86,15 %	54,23 %	84,26 %	63,27 %	73,76 %	71,43 %	84,26 %	63,99 %	62,24%	72,16%	55,25%	62,83%
	PRECION	95,54 %	93,77 %	92,64 %	92,84 %	95,79 %	95,91 %	95,40 %	95,69 %	93,03 %	93,53 %	92,55 %	95,60%	95,99%	95,74%	92,83%
13	RECALL	92,94 %	96,07 %	78,56 %	87,92 %	61,74 %	89,15 %	73,13 %	81,41 %	76,12 %	84,40 %	69,06 %	70,69%	78,02%	64,04%	66,76%
	PRECION	95,90 %	95,08 %	95,24 %	93,89 %	96,59 %	96,90 %	96,10 %	95,98 %	95,75 %	94,83 %	94,98 %	96,67%	96,81%	96,12%	95,49%
	RECALL	68,02 %	84,22 %	72,69 %	89,30 %	41,23 %	58,36 %	51,19 %	61,99 %	63,24 %	76,12 %	64,80 %	45,17%	53,58%	46,31%	57,11%
15	PRECION	98,51 %	97,27 %	97,85 %	97,36 %	99,38 %	98,39 %	99,19 %	98,57 %	97,74 %	97,16 %	97,87 %	99,15%	98,46%	99,41%	97,76%
	RECALL	80,98 %	92,79 %	69,94 %	90,64 %	49,39 %	75,15 %	56,60 %	74,08 %	66,26 %	84,05 %	63,34 %	53,99%	68,71%	51,69%	60,12%
	PRECION	99,06 %	99,09 %	99,00 %	98,67 %	99,26 %	99,25 %	99,30 %	99,09 %	99,06 %	99,15 %	99,09 %	99,35%	99,24%	99,30%	99,09%
16	RECALL	66,09 %	87,35 %	71,09 %	89,72 %	42,00 %	59,16 %	50,98 %	58,40 %	66,13 %	78,11 %	62,96 %	47,90%	52,23%	44,59%	58,58%
	PRECION	98,28 %	93,75 %	93,12 %	92,15 %	98,13 %	98,25 %	97,79 %	98,57 %	94,25 %	93,99 %	93,55 %	97,77%	98,54%	98,15%	94,87%
	RECALL	65,24 %	94,02 %	73,22 %	90,31 %	44,73 %	64,10 %	50,43 %	58,98 %	70,09 %	84,62 %	66,10 %	49,86%	57,83%	45,30%	63,25%
18	PRECION	92,95 %	91,75 %	92,23 %	91,69 %	93,91 %	92,78 %	93,82 %	93,08 %	92,60 %	91,89 %	92,55 %	93,68%	92,88%	94,07%	92,79%
	RECALL	85,05 %	92,03 %	76,56 %	91,27 %	53,79 %	79,89 %	63,35 %	77,62 %	69,35 %	84,29 %	69,73 %	58,50%	73,29%	57,81%	63,51%
	PRECION	98,64 %	97,93 %	98,44 %	98,02 %	98,47 %	98,54 %	98,69 %	98,50 %	98,34 %	97,87 %	98,29 %	98,59%	98,39%	98,57%	98,18%
19	RECALL	71,35 %	93,12 %	70,93 %	90,45 %	45,08 %	66,57 %	52,81 %	64,33 %	66,43 %	83,99 %	64,75 %	49,16%	59,97%	48,46%	60,53%
	PRECION	99,31 %	95,15 %	96,58 %	95,39 %	99,38 %	99,28 %	99,30 %	99,39 %	96,59 %	95,42 %	96,83 %	99,27%	99,35%	99,40%	96,87%
	RECALL	72,13 %	96,10 %	76,12 %	88,91 %	48,15 %	69,33 %	56,34 %	63,84 %	73,53 %	85,41 %	67,03 %	54,55%	61,44%	49,55%	64,84%
21	PRECION	93,40 %	93,56 %	92,87 %	92,77 %	95,37 %	94,28 %	93,69 %	93,96 %	93,82 %	94,09 %	93,40 %	94,43%	94,86%	94,43%	94,59%
	RECALL	80,66 %	93,70 %	66,42 %	86,51 %	46,33 %	76,61 %	55,62 %	70,01 %	63,72 %	81,11 %	57,27 %	53,37%	66,42%	48,28%	55,02%

		97,36 %	95,33 %	94,02 %	93,52 %	97,45 %	97,42 %	97,27 %	97,54 %	95,40 %	95,47 %	93,93 %	97,29%	97,64%	97,42%	95,60%
22	PRECISION															
	RECALL	91,33 %	92,67 %	79,46 %	84,83 %	59,13 %	85,86 %	73,58 %	77,61 %	74,92 %	78,22 %	67,08 %	70,28%	72,65%	62,23%	62,85%
		91,67 %	81,82 %	81,82 %	81,58 %	100,00 %	91,43 %	100,00 %	90,01 %	83,87 %	83,33 %	85,19 %	100,00%	89,66%	100,00 %	88,00%
23	PRECISION															
	RECALL	89,19 %	97,30 %	72,97 %	83,78 %	51,35 %	86,49 %	64,86 %	72,97 %	70,27 %	81,08 %	62,16 %	62,16%	70,27%	54,05%	59,46%
		97,43 %	94,60 %	94,12 %	94,19 %	97,69 %	97,43 %	97,89 %	97,43 %	94,67 %	94,42 %	93,87 %	97,83%	97,24%	97,73%	94,55%
24	PRECISION															
	RECALL	75,96 %	81,05 %	81,36 %	92,45 %	52,08 %	64,25 %	64,25 %	70,11 %	68,41 %	75,65 %	75,50 %	55,62%	59,78%	59,63%	64,10%
	MÉDIA PRECISION	96,03 %	91,86 %	92,16 %	91,17 %	96,46 %	96,14 %	96,35 %	96,04 %	92,66 %	91,93 %	92,24 %	96,44%	96,05%	96,37%	92,84%
	MÉDIA RECALL	81,33 %	91,77 %	74,35 %	88,92 %	51,24 %	75,55 %	61,65 %	72,04 %	69,24 %	81,52 %	66,18 %	57,86%	66,80%	54,72%	61,56%

Com base nos valores das médias destacados nas duas últimas linhas da Tabela 30, é possível observar que, quando consideramos a rejeição de estudos não ranqueados, os critérios possuem precisão alta e um valor de *recall* considerável. A média de precisão dos estudos foi de 96,46% para o critério OU (combinação de todos os critérios individuais propostos) com *recall* de 51,24% para esse mesmo critério, como destacado nas duas últimas linhas da tabela.

A Tabela 31 apresenta a penúltima linha da Tabela 30 (médias calculadas de *precision*) reordenada em ordem decrescente pelos valores de média. A precisão permite avaliar os critérios com relação a quanto acertam. O critério com maior precisão é aquele que menos rejeita estudos relevantes de forma automática.

A Tabela 32 apresenta a última linha da Tabela 30 (médias calculadas de *recall*) reordenada em ordem decrescente pelos valores de média. Quanto maior o valor de *recall*, maior a quantidade de estudos classificados como rejeitados automaticamente, diminuindo o esforço do pesquisador no que diz respeito à avaliação dos estudos restantes.

Tabela 31. Critérios classificados pelo valor de *precision* (do maior para o menor)

OU	SC+CIT+AU	SC+AU+LO	SC+AU	SC + CIT	SC+CIT+LO	SC+LO	SC	CIT+AU+LO	CIT + AU	AU+LO	AU	CIT+LO	CIT	LO
96,46 %	96,44%	96,37%	96,35 %	96,14%	96,05%	96,04%	96,03%	92,84%	92,66%	92,24 %	92,16 %	91,93 %	91,86 %	91,17 %

Tabela 32. Critérios classificados pelos valores de *recall* (do maior para o menor)

CIT	LO	CIT+LO	SC	SC + CIT	AU	SC+LO	CIT + AU	SC+CIT+LO	AU+LO	SC+AU	CIT+AU+LO	SC+CIT+AU	SC+AU+LO	OU
91,77 %	88,92 %	81,52 %	81,33 %	75,55%	74,35 %	72,04 %	69,24%	66,80%	66,18 %	61,65 %	61,56%	57,86%	54,72%	51,24 %

O melhor critério para rejeitar estudos é aquele que tiver a maior precisão e o maior *recall* possíveis. Se o pesquisador quer encontrar o maior número de estudos

relevantes, adota-se o critério de maior precisão. Caso o pesquisador deseje reduzir ao máximo o esforço, classificando o maior número possível de estudos como rejeitados automaticamente, porém assumindo uma possível perda de estudos relevantes, adota-se o critério com maior *recall*. Um pesquisador poderá utilizar a ordem de leitura de referências aplicando os critérios apresentadas na Tabela 31, caso deseje uma maior precisão. Caso o pesquisador prefira um maior *recall*, poderá seguir a ordem de leitura de referências aplicando os critérios apresentados na Tabela 32.

Para utilizar os critérios de priorização propostos para ranquear os estudos, o pesquisador poderia adotar, com base nos resultados obtidos neste experimento, a seguinte ordem de leitura:

1. Ranquear pelo critério SC e avaliar manualmente os estudos priorizados por ele;
2. Ranquear pelo critério AU e avaliar manualmente os estudos priorizados por ele;
3. Ranquear pelo critério CIT e avaliar manualmente os estudos priorizados por ele;
4. Ranquear pelo critério LO e avaliar manualmente os estudos priorizados por ele;
5. Ranquear, em seguida, pelo critério OU e avaliar manualmente os estudos priorizados por ele;
6. Ranquear os demais estudos.

A ordem apresentada lista os critérios em ordem de precisão. O sexto item determina a leitura de todos os estudos não ranqueados. Um pesquisador que utilize os critérios poderá parar a leitura em qualquer um dos passos anteriores. Quanto mais passos ele executar, mais preciso será o método e maior o esforço envolvido. Quando o pesquisador decidir finalizar a pesquisa, a precisão do método será equivalente à precisão do critério utilizado. Por exemplo, caso o pesquisador encerre a avaliação de referências no passo 2. A precisão envolvida ao rejeitar os demais estudos será a combinação dos critérios SC+AU (96,35%) e o número de referências relevantes perdidas é de 3,65% do número de estudos encontrados (relevantes ou não). Caso o pesquisador utilize a classificação manual nos estudos ranqueados até o passo 5, a precisão será do Critério OU de (96,46%)

Ressalta-se que essa lista foi construída com base nos resultados do experimento realizado. Em outro contexto a sequência de critérios pode ser diferente. Mais estudos são necessários para avaliar a efetividade e a melhor ordem dos critérios em diferentes cenários. Os valores de precisão apresentados não podem ser generalizados.

O número de estudos que podem ser filtrados na revisão é considerável (menor valor médio obtido no estudo foi 51,24%). Aplicar os critérios pode reduzir o volume de referências a serem processadas caso o pesquisador opte por eliminar os estudos não ranqueados. Vale ressaltar que, ao aplicar o critério para reduzir o volume de referências, existe o risco de descartar algumas que sejam relevantes.

A Tabela 33 mostra as porcentagens de rejeição e de erro (estudos aceitos que serão rejeitados posteriormente) caso os critérios sejam aplicados para rejeição automática. A coluna P representa o participante e a coluna GR representa o grupo. As colunas 3 e 4 (Rejeitados e Aceitos) indicam, respectivamente, o número de estudos que foram considerados não-relevantes ou relevantes após verificação manual. Com base na tabela, é possível observar que uma grande quantidade de referências não precisa ser avaliada pelo pesquisador que utiliza os critérios (vide colunas Rejeitados de cada critério utilizado), contudo há um percentual de erro gerado, conforme mostrado nas colunas “Erro” da tabela.

Pode-se interpretar os dados da Tabela 33 da seguinte forma: se a porcentagem de rejeição do critério for maior que a porcentagem de erro, o critério auxilia na rejeição dos estudos e o conjunto de estudos que sobra para ser avaliado manualmente possui uma taxa de aceitação maior de seus elementos do que o conjunto original com todos os elementos.

Nota-se que os critérios OU, SC, LO e AU, na maioria dos casos, se mostram efetivos. O pior caso apresentado pelos critérios pode ser avaliado na linha de dados do participante (8). Nesse caso em particular, a taxa de erro foi extremamente alta, e o algoritmo descartou mais informação relevante do que de fato ajudou o participante a reduzir o esforço empregado. Em contrapartida, existem exemplos com bom desempenho. É o caso dos participantes (6) e (23) em relação ao critério OU, que não teve erro, sendo esses casos os considerados ótimos com relação à precisão do algoritmo.

Tabela 33. Efetividade dos critérios de priorização para rejeição

Revisor Manual				Critério SC				Critério AU				Critério CIT				Critério LO				Critério OU			
P	Gr	Rejeitados	Aceitos	Rejeitados		Erro		Rejeitados		Erro		Rejeitados		Erro		Rejeitados		Erro		Rejeitados		Erro	
				Nº	(%)	Nº	(%)	Nº	(%)	Nº	(%)	Nº	(%)	Nº	(%)	Nº	(%)	Nº	(%)	Nº	(%)	Nº	(%)
1	CBS	1254	108	1096	80%	36	33%	1032	76%	56	52%	1208	89%	81	75%	1192	88%	90	83%	716	53%	15	14%
2	EP	747	26	717	93%	17	65%	574	74%	22	85%	754	98%	26	100%	664	86%	22	85%	442	57%	11	42%
3	CES	326	41	317	86%	10	24%	273	74%	23	56%	287	78%	26	63%	339	92%	36	88%	189	51%	4	10%
4	EQ	305	22	255	78%	12	55%	246	75%	18	82%	321	98%	22	100%	284	87%	18	82%	162	50%	8	36%
5	CES	990	10	861	86%	2	20%	680	68%	3	30%	933	93%	10	100%	879	88%	9	90%	494	49%	1	10%
6	CBS	169	13	118	65%	0	0%	146	80%	4	31%	170	93%	9	69%	162	89%	11	85%	88	48%	0	0%
7	HE	95	9	97	93%	6	67%	75	72%	7	78%	102	98%	8	89%	89	86%	9	100%	58	56%	4	44%
8	CC	651	36	651	95%	31	86%	551	80%	32	89%	596	87%	31	86%	618	90%	35	97%	395	57%	24	67%
9	CES	13	11	10	42%	1	9%	15	63%	6	55%	24	100%	11	100%	23	96%	11	100%	7	29%	1	9%
10	EP	1243	34	940	74%	13	38%	928	73%	13	38%	1000	78%	13	38%	1156	91%	30	88%	557	44%	1	3%
11	CES	163	37	151	76%	15	41%	153	77%	24	65%	169	85%	29	78%	186	93%	32	86%	96	48%	9	24%
12	EQ	686	116	638	80%	46	40%	588	73%	90	78%	784	98%	115	99%	692	86%	101	87%	403	50%	31	27%
13	EP	737	56	717	90%	32	57%	625	79%	46	82%	755	95%	47	84%	698	88%	50	89%	475	60%	20	36%
14	CC	964	60	683	67%	28	47%	735	72%	35	58%	853	83%	42	70%	916	89%	56	93%	411	40%	14	23%
15	CES	653	17	536	80%	8	47%	466	70%	10	59%	622	93%	17	100%	607	91%	16	94%	324	48%	2	12%
16	CBS	2238	32	1493	66%	14	44%	1607	71%	16	50%	1973	87%	18	56%	2035	90%	27	84%	947	42%	7	22%
17	CIA	351	30	233	61%	4	13%	276	72%	19	63%	352	92%	22	73%	344	90%	27	90%	160	42%	3	10%
18	CES	1318	123	1206	84%	85	69%	1094	76%	85	69%	1322	92%	109	89%	1312	91%	109	89%	755	52%	46	37%
19	CES	713	14	515	71%	7	50%	513	71%	8	57%	678	93%	14	100%	644	89%	13	93%	326	45%	5	36%
20	EP	1001	52	727	69%	5	10%	789	75%	27	52%	1011	96%	49	94%	933	89%	43	83%	485	46%	3	6%
21	EQ	667	55	576	80%	38	69%	477	66%	34	62%	668	93%	43	78%	622	86%	45	82%	324	45%	15	27%
22	EP	969	65	909	88%	24	37%	819	79%	49	75%	942	91%	44	68%	879	85%	57	88%	588	57%	15	23%
23	CIA	37	9	36	78%	3	33%	33	72%	6	67%	44	96%	8	89%	38	83%	7	78%	19	41%	0	0%
24	CC	649	39	506	74%	13	33%	561	82%	33	85%	556	81%	30	77%	637	93%	37	95%	346	50%	8	21%

A Tabela 34 apresenta as médias dos valores de estudos rejeitados automaticamente (coluna Rejeitados) e as médias de erros (coluna Erro) que representam o número de estudos que seriam aceitos na revisão manual, mas foram rejeitados por não estarem ranqueados por nenhum dos critérios propostos. Com base nos dados apresentados, pode-se observar que o critério CIT rejeita o maior número de estudos (91%) e o critério OU tem a menor taxa de erros (22%), o que é um valor que possivelmente possa ser melhorado.

Tabela 34. Média dos valores de rejeição automática e erros para os critérios de priorização

Critério SC		Critério AU		Critério CIT		Critério LO		Critério OU	
Rejeitados	Erro	Rejeitados	Erro	Rejeitados	Erro	Rejeitados	Erro	Rejeitados	Erro
77%	41%	74%	63%	91%	82%	89%	89%	48%	22%

Por não ter sido realizada uma avaliação qualitativa neste estudo, não é possível inferir a relevância das novas referências encontradas para as pesquisas. Entretanto, ressalta-se que quantitativamente, há indícios de que os critérios são favoráveis para a redução de esforço na aplicação de *snowballing*.

Em resumo, respondendo as questões levantadas no planejamento do experimento, podemos concluir que:

- Com base nos resultados de kappa apresentado na Tabela 29, não se pode afirmar que critérios de priorização podem ser utilizados para a classificação automática de todo o conjunto de estudos encontrados devido ao grande volume de falsos negativos encontrados. Os resultados mostram um baixo nível de concordância entre os critérios e o revisor manual.
- Os critérios de priorização mostraram-se úteis para a redução do volume a ser avaliado. Esse resultado é concluído com base nos dados das Tabelas 30 e 31. Ressalta-se que o principal problema é a possível perda de estudos relevantes, que deve ser avaliada.
- Sugere-se a aplicação dos critérios de priorização, com base nos resultados apresentados nas Tabelas 30 e 31, na seguinte ordem: SC, AU, CIT e LO, O critério OU é a combinação dos quatro critérios e é equivalente a aplicar os quatro em sequência.

-
- *Snowballing backward* melhora os resultados de busca com *strings*, com base nos dados apresentados na Tabela 24. Esse resultado já havia sido concluído em um estudo anterior e está presente na literatura, mas foi reforçado.
 - Com base nos dados apresentados na Tabela 27, é possível encontrar estudos relevantes aplicando *snowballing* em estudos rejeitados pois, em diversos casos apresentados no estudo, *snowballing* aplicado a estudos rejeitados permitiu encontrar novos estudos relevantes para a pesquisa.
 - O uso do suporte computacional mostrou-se positivo. Nas Tabela 25 e 26 são apresentados dados referentes às duplicações que foram identificadas de forma automática. Na Figura 17 são apresentados dados referentes à percepção dos participantes em relação ao apoio computacional utilizado, que mostrou-se muito positivo.

5.5 Considerações Finais

Realizou-se um conjunto de estudos como parte da avaliação deste trabalho. Diversas dificuldades foram encontradas na condução desses estudos, dentre elas, a seleção dos participantes e o tempo gasto por eles para executar as tarefas propostas de forma abrangente.

Foi utilizado como conjunto inicial para *snowballing* os resultados obtidos pela busca com *strings*. A mesma mostrou-se efetiva. Não foi possível testar outros conjuntos iniciais e é possível que, com outros conjuntos, os resultados sejam diferentes.

Os algoritmos da ferramenta StArt foram avaliados e os resultados mostram indícios de que o apoio computacional poderá auxiliar os pesquisadores que executam *snowballing*.

Os critérios de priorização mostraram-se efetivos. Uma ordem foi proposta, mas ainda são necessárias novas pesquisas para avaliar os critérios em diferentes contextos.

CONCLUSÃO

Este capítulo descreve as conclusões deste trabalho, as contribuições e limitações existentes, as lições aprendidas ao longo da pesquisa, as oportunidades futuras que este trabalho de pesquisa oferece e as publicações feitas pelo autor deste trabalho.

6.1 Conclusões

Neste trabalho foram definidos critérios para priorização de estudos primários identificados por *snowballing* com conjunto inicial gerado por *string* de busca. Como apresentado no Capítulo 1, o objetivo geral era propor critérios de priorização para tornar a execução do *snowballing* mais eficiente do que a execução do *snowballing* completo, porém mantendo a qualidade dos resultados, isto é, perdendo o menor número possível de estudos primários relevantes referentes ao tópico da pesquisa.

O contexto que motivou esse trabalho foi a pouca quantidade de estudos publicados sobre diretrizes completas de como executar *snowballing*, além do esforço excessivo que pode ser requerido ao executar *snowballing* utilizando estratégias e pesquisas abrangentes, que podem levar a milhares de referências para serem manualmente avaliadas pelo pesquisador. Nesse sentido, o suporte computacional a *snowballing*, também não facilmente encontrado na literatura, pode ser de grande contribuição para o pesquisador, o que motivou a implementar o módulo de suporte a *snowballing* na ferramenta StArt.

Vale ressaltar que o conjunto inicial de estudos primários relevantes, neste trabalho, os quais permitem a aplicação dos critérios de priorização propostos, foi fixado como sendo o resultado de buscas com *string*, não sendo possível, por questões de tempo, explorar outras estratégias para composição do *seed* para aplicação de *snowballing*.

O trabalho foi avaliado com base em três estudos experimentais. O primeiro foi um experimento realizado para avaliar as dificuldades de aplicação de *snowballing backward* e *forward* e identificar pontos de melhorias na aplicação da técnica, auxiliando também na elaboração dos critérios de priorização. O segundo foi

um estudo exploratório para avaliar os algoritmos de extração de referências e de similaridade propostos visto que esses algoritmos são fundamentais para a aplicação dos critérios com suporte computacional da ferramenta StArt. O terceiro foi um experimento para avaliar a aplicação dos critérios de priorização de estudos primários propostos, que é a principal contribuição deste trabalho.

Como resultado do primeiro estudo experimental, temos que *snowballing backward* ou *forward* permitem identificar novos estudos relevantes para *strings* de busca. *Snowballing backward*, no contexto do estudo realizado, localizou mais estudos relevantes do que *snowballing forward*.

Como resultado do segundo estudo experimental, temos que os algoritmos desenvolvidos na StArt se mostram promissores para uma abordagem inicial. No estudo realizado, foi possível reconhecer em média 87% das referências automaticamente detectadas pela ferramenta. O algoritmo de similaridade teve 100% de precisão para identificar duplicações no conjunto de referências estudos. Ressaltando que não é possível generalizar esse resultado e que o algoritmo é passível de erros.

Como resultado do terceiro estudo experimental, temos que os critérios de priorização de referências apresentam bom resultado. O critério mais preciso foi o critério OU (combinação dos quatro critérios) com *precision* de 96,46% e *recall* de 51,24%.

Sendo assim, com a condução desses estudos, considera-se ter obtido subsídios de que os critérios de priorização propostos, que fazem uso dos algoritmos de extração de referências e de similaridade desenvolvidos, com o apoio do módulo criado para tal fim na ferramenta StArt, contribuem para o aumento da eficiência na aplicação da técnica *snowballing* sem que haja perda significativa de qualidade dos resultados obtidos.

6.2 Contribuições e limitações da pesquisa

As seguintes contribuições desta pesquisa podem ser destacadas:

- 1) Proposição de critérios de priorização de estudos primários identificados por *snowballing* com conjunto inicial gerado por *string* de busca.

-
- 2) Sugestão de estratégias de aplicação distintas de *snowballing* com base nos critérios de priorização propostos, tendo em vista o interesse do pesquisador em obter maior precisão de resultados utilizando maior esforço, ou em aplicar o *snowballing* seguindo as diretrizes convencionais, com esforço menor, porém com precisão menor de resultados.
 - 3) Desenvolvimento de apoio computacional a *snowballing* por meio de:
 - a. Desenvolvimento de uma estratégia baseada em expressões regulares para reconhecimento e extração de referências dentro de um texto completo de estudos (PDF) ou de uma lista de referências (BibTex).
 - b. Desenvolvimento de um algoritmo para comparação de cadeia de caracteres para identificar duplicidade de estudos e citações de estudos, de forma mais efetiva que o proposto por Salton e Allan, 1994) para comparação de referências com ruído.
 - c. Evolução da ferramenta StArt com a implementação do módulo para aplicação de *snowballing*.
 - d. Metáfora visual por meio de grafos e nuvens de palavras para aplicação de *snowballing*, implementada na ferramenta StArt
 - e. A versão beta da StArt que inclui alguns recursos para *snowballing* já está disponível para a comunidade desde 21/06/2016 e vem recebendo constantes atualizações. Hoje já conta com 1300 *downloads* em um período de seis meses, o que pode demonstrar interesse da comunidade sobre esse tipo de pesquisa, embora a versão conte com outras funcionalidades novas.
 - 4) Divulgação de resultados obtidos por meio de um mapeamento sistemático avaliando o uso prático de *snowballing* pela comunidade e de uma revisão sistemática para conhecer as estratégias propostas na literatura para aplicação da técnica.

Em relação às limitações do trabalho, podemos destacar:

- 1) Apenas buscas com *strings* foram consideradas para a composição do conjunto inicial de *snowballing*. Não foram investigadas outras estratégias para composição do *seed*, o que pode levar a resultados diferentes ao aplicar os critérios de priorização.

-
- 2) Um único estudo exploratório foi realizado para a avaliação dos algoritmos de extração de referências e de similaridade para comparação de dados, implementados na ferramenta StArt. Para maior confiabilidade nos algoritmos implementados, novos estudos devem ser conduzidos.
 - 3) As estratégias para aplicação de *snowballing* utilizando os critérios de priorização de estudos primários definidos foram propostas com base nos resultados de apenas um experimento. Elas devem ser mais bem avaliadas com a realização de novos estudos experimentais.
 - 4) Não foram realizados estudos que avaliassem qualitativamente as publicações encontradas ou descartadas por cada estratégia de aplicação dos critérios. Para definir, ou comparar as diferentes estratégias, essa informação faz-se necessária.
 - 5) Como apenas as estratégias para recuperar a lista de referências de estudos por meio do texto completo (PDF) e de arquivos de referências (BibTex) foram implementadas, só é possível executar *snowballing backward* com apoio automatizado de extração de referências. *Snowballing forward* só pode ser executado na ferramenta com a inserção manual das referências, pois não é possível obtê-las por meio do PDF ou BibTex nesse caso.
 - 6) Nos experimentos realizados, foram selecionados como participantes apenas alunos de pós-graduação (mestrado ou doutorado). Assim os resultados podem ser diferentes para pesquisas realizadas por pesquisadores experientes.

6.3 Lições aprendidas

Avaliar métodos de busca é uma atividade complexa. É complicado definir o quão preciso é um método de busca sem conhecer os cenários no qual é aplicado. É necessária a realização de análises quantitativas e qualitativas, sendo essa uma tarefa bastante complexa.

As estratégias de priorização de referências foram avaliadas sempre em caráter exploratório, os critérios de priorização foram testados e os próprios

resultados foram utilizados para melhorar as estratégias. Considera-se essa abordagem assertiva, mas é ressaltado que novos estudos devem ser executados, uma vez que as estratégias estejam finalizadas.

Uma das principais dificuldades para a condução dos estudos deste trabalho foi a limitação de tempo. Não foi explorar todas as estratégias com os participantes, e nem foi possível aplicar *snowballing* de forma ilimitada dado o grande esforço que o estudo consumiria dos participantes, contudo foram obtidos resultados que apresentam indícios de uma boa efetividade dos critérios de priorização em cenários reais de pesquisa.

Outra dificuldade foi estimar e ter de trabalhar com o erro envolvido ao aplicar os critérios para priorização de estudos e os algoritmos para extração de referências e cálculo de similaridade implementados na ferramenta StArt. Esses erros limitaram possíveis conclusões dos estudos.

6.4 Oportunidades Futuras

A seguir, apresentam-se algumas ideias para continuidade desta pesquisa:

- Condução de outros estudos experimentais a fim de explorar todas as possibilidades e técnicas para aplicação de *snowballing* e reforçar os resultados já obtidos nos estudos apresentados nesse trabalho.
- Proposição e avaliação de novos critérios de priorização de referências para complementar os critérios já propostos. Entre as possibilidades está o uso de critérios semi-supervisionados ou supervisionados pelo pesquisador com base nas classificações da própria revisão.
- Realizar uma avaliação qualitativa dos estudos encontrados por *snowballing*. Existem poucos estudos que avaliam de forma qualitativa as buscas realizadas e, antes de afirmar se o método é muito ou pouco efetivo para encontrar novos estudos relevantes, esse tipo de análise faz-se necessária.
- Estudar uma estratégia conjunta entre SCAS e *snowballing* para a classificação semiautomática de referências.
- Avaliar as metáforas visuais (grafo de referências das referências, e

nuvem de palavras) propostas na ferramenta StArt para *snowballing*.

- Melhorar os algoritmos de reconhecimento de referências e de comparação de referências na ferramenta StArt.
- Desenvolver outras estratégias para a coleta de referências e citações na ferramenta StArt que viabilizem melhores resultados.
- Geração de tutoriais em vídeo na ferramenta StArt para aplicação de *snowballing* com base nos critérios de priorização propostos.

6.5 Publicações

Nesta seção são apresentadas as publicações do autor desta pesquisa durante o período de mestrado, explicitando sua contribuição nas publicações. As publicações estão organizadas em dois tópicos: capítulos de livros e artigos completos publicados em anais de congressos (*full papers*). Duas das publicações são anteriores ao trabalho de mestrado sendo do ano de 2012, mas foram evidenciadas por terem inspirado este trabalho de mestrado.

- **Capítulos de livros publicados:**

- 1) FABBRI, S. C. P. F.; HERNANDES, E. C. M.; DI THOMMAZO, A.; BELGAMO, A.; ZAMBONI, A.; SILVA, C. Using information visualization and text mining to facilitate the conduction of systematic literature reviews. In: José Cordeiro, Leszek A. Maciaszek, Joaquim Filipe. (Org.). *Lecture Notes in Business Information Processing*. 1ed. Berlin: Springer, 2013, v. 141, p. 1-12.

Essa publicação é anterior, mas serviu de inspiração para esse trabalho. É uma extensão de um artigo anterior publicado em congresso.

- **Trabalhos completos publicados em anais de congressos:**

- 1) FABBRI, S. C. P. F.; HERNANDES, E. C. M.; THOMMAZO, A.; BELGAMO, A.; ZAMBONI, A.; SILVA, C. Managing literature reviews

information through visualization. In: *International Conference on Enterprise Information Systems, 2012, Wroclaw. International Conference on Enterprise Information Systems, 2012. v. 1. p. 1-10.*

Essa publicação é anterior a este trabalho, mas serviu de inspiração para esse projeto. Foi fruto do envolvimento do autor desta pesquisa com o projeto StArt. Esse trabalho recebeu o prêmio de melhor artigo (*best paper*) da conferência.

- 2) GASTALDI, R. A.; SILVA, C.; DI THOMMAZO, A.; HERNANDES, E.; BITTENCOURT, D.; FABBRI, S. Insight: uma ferramenta para análise qualitativa de dados apoiada por mineração de texto e visualização de informações. In: *Brazilian Conference on Software: Theory and Practice (Tools Session), 2015, Belo Horizonte. Proceedings of Brazilian Conference on Software: Theory and Practice (Tools Session). Belo Horizonte: Universidade Federal de Minas Gerais, 2015. v. 1. p. 129-136.*

Essa publicação foi fruto do envolvimento do autor desta pesquisa com o projeto da ferramenta *Insight* desenvolvida pelo grupo de pesquisa e associada com a ferramenta StArt.

- 3) FABBRI, S., OCTAVIANO, F., SILVA, C., DI THOMMAZO, A., BELGAMO, A., HERNANDES, E. Improvements in the StArt tool to better support the systematic review process. In *20th International Conference on Evaluation and Assessment in Software Engineering. 2016, Limerick, Ireland,*

Essa publicação relata o apoio computacional da Ferramenta StArt, incluindo os resultados preliminar desenvolvido nesse trabalho.

- 4) OCTAVIANO, F., SILVA C., FABBRI S. Using the SCAS strategy to perform the initial selection of studies in systematic reviews: an experimental study. In *20th International Conference on Evaluation and Assessment in Software Engineering. 2016, Limerick, Ireland.*

Esse trabalho é fruto de uma colaboração do autor dentro do grupo de pesquisa. A implementação do SCAS utiliza o número de citação gerado pelos algoritmos desenvolvidos neste trabalho.

REFERÊNCIAS

ABNT. **ABNT NBR 15606-2:2007-** Associação Brasileira de Normas Técnicas, Rio de Janeiro/RJ- Brasil, 2007.

ACM DL, **ACM Digital Library**, 2017, Disponível em <<http://dl.acm.org/>>. Acesso em: 15 jan. de 2017.

AL-ZUBIDY, A; CARVER, J. C.; Review of Systematic Literature Review Tools. **Technical Report SERG-2014-03**. Department of Computer Science, The University of Alabama, Alabama, USA, 2014.

BADAMPUDI D.; WOHLIN C., PETERSEN K. Experiences from using snowballing and database searches in systematic literature studies, In: **Proceedings of 19th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, Nanjing, China: ACM, p.1-10, 2015.

BASILI, V. R. Software modeling and measurement: The goal/question/metric paradigm. **Technical Report**, University of Maryland at College Park, College Park, MD, USA, 1992.

BEZERRA, F.; FAVACHO C. H.; SOUZA, R.; de SOUZA, C. Towards Supporting Systematic Mappings Studies: Automatic Snowballing Approach In: **Proceedings of 29th Simpósio Brasileiro de Banco de Dados (SBBd)**, Curitiba, Brasil: JIDM. p.167-176, 2014.

BibTex, **Padrão de arquivos BibTex**, 2017, Disponível em <<http://www.bibtex.org/>>. Acesso em: 15 jan. 2017.

BING, **Motor de pesquisa da Microsoft**, 2017, Disponível em <<https://br.bing.com/>>. Acesso em: 15 jan. 2017.

BIOLCHINI, J., MIAN, P. G., NATALI, A. C.; TRAVASSOS, G. H. Systematic review in software engineering: Relevance and utility. **Technical Report RT-ES**, Programa de Engenharia de Sistemas e Computação. Universidade Federal do Rio de Janeiro, Rio de Janeiro, BR, 2005.

BRERETON, P.; KITCHENHAM, B. A.; BUDGEN, D.; TURNER, M.; KHALIL, M. Lessons from applying the systematic literature review process within the software engineering domain. **Systems and Software (JSS)**, v.80, p. 571-583, 2007.

CARLETTA, J. Assessing agreement on classification tasks: The kappa statistic. **Computational Linguistics (IJCL)**, v.22, n.2, p 249–254, 1996.

CHOONG, M. K.; GALGANI, F.; DUNN, A. G.; TSAFNAT, G. Automatic Evidence Retrieval for Systematic Reviews, **Medical Internet Research(JMIR)**, v.16, p. 16-26 , 2014.

DA SILVA, F. Q. B.; SANTOS, A. L. M.; SOARES, S.; FRANÇA, A. C. C.; MONTEIRO, C V.F.; MACIEL, F. F. Six Years of Systematic Literature Reviews in Software Engineering: An Updated Tertiary Study. **Information and Software Technology (IST)**, v.53, n.9, p. 899-913, 2011.

DELUCA J. B., MULILINS, M. M.; LYLES, M. C.; CREPAZ N. Developing a Comprehensive Search Strategy for Evidence Based Systematic Reviews. **Evidence Based Library and Information Practice (EBLIP)**, v. 3 n. 1, p. 3-32, 2008.

DICKERSIN, K.; SCHERER, R.; LEFEBVRE, C. Systematic Reviews: Identifying Relevant Studies for Systematic Reviews. **British Medical Journal (BMJ)**, v. 309, n. 6964, p. 1286-1291, 1994.

DIESTE, O.; PADUA, A. G. Developing Search Strategies for Detecting Relevant Experiments for Systematic Reviews. **Empirical Software Engineering (EMSE)**, v. 14, p. 513 - 539, 2014.

DYBÅ, T.; DINGSØYR, T. Strength of evidence in systematic reviews in software engineering. In: **Proceedings of the 2th ACM-IEEE international symposium on Empirical software engineering and measurement (ESEM)**, Kaiserslautern, Germany: ACM, p. 178-187, 2008.

FABBRI, S.; FELIZARDO, K. R.; FERRARI, F.; HERNANDES, E.; OCTAVIANO, F.; NAKAGAWA, E.; MALDONADO, J. C. Externalising tacit knowledge of the systematic review process. **IET Software (IET)**, v. 7, n. 6, p. 298-307, 2013.

FABBRI, S.; HERNANDES, E.; DI THOMMAZO, A.; BELGAMO, A., ZAMBONI, A.; SILVA, C. Using information visualization and text mining to facilitate the conduction of systematic literature reviews. In: **Proceedings of 14th International**

Conference on Enterprise Information (ICEIS), Wroclaw, Poland: Springer, pp. 243-256, 2012.

FELIZARDO, K. R.; MENDES E.; KALINOWSKI M.; SOUZA E. F.; VIJAYKUMAR N. L. Using Forward Snowballing to update Systematic Reviews in Software Engineering. In: **Proceedings of 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)**, Ciudad Real, Spain: ACM, p 1-6, 2016.

GAROUSI, V.; FELDERER, M.; MÄNTYLÄ, M. V. The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. In: **Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, Limerick, Ireland: ACM, p.1-6, 2016.

GOOGLE SEARCH, **Serviço de pesquisas online da Google**, 2017, Disponível em <<https://www.google.com.br/>>. Acesso em: 15 jan. 2017.

GREENHALGH, T. How to read a paper: The basics of evidence-based medicine. 15^o edição. **BMJ Books e Wiley**, 2014.

GUIMARÃES, A. Structured Abstracts. Narrative Review. **Acta Cirurgica Brasileira**; São Paulo, Brasil, v.21, n. 4, pp. 263-268, 2006.

HERNANDES E.; BELGAMO A.; FABBRI S.; Experimental Studies in Software Inspection Process: *A Systematic Mapping*. In: **Proceedings of the 15th International Conference on Enterprise Information Systems (ICEIS)**, Angers, France: Springer, p. 66-76, 2013.

HERNANDES, E.; ZAMBONI, A.; FABBRI, S.; DI THOMMAZO, A. Using GQM and TAM to evaluate StArt - a tool that supports Systematic Review. **CLEI Eletronic (CLEI)**, v.15, n.1, 2012.

IEEE Xplore, **IEEE Xplore Digital Library**, 2017, Disponível em <<http://ieeexplore.ieee.org/>>. Acesso em: 17 ago. 2017.

JALALI, S.; WOHLIN, C. Systematic Literature Studies: Database Searches vs. Backward Snowballing In: **Proceedings of ACM-IEEE international symposium on Empirical software engineering and measurement (ESEM)**. Lund, Sweden: ACM-IEEE, p. 29-38, 2012;

JEDLITSCHKA, A.; CIOLKOWSKI, M. Towards evidence in software engineering. **Empirical Software Engineering (JESE)**, Washington, USA, p. 261-270, 2004.

KITCHENHAM, B. Procedures for performing systematic reviews. **Technical Report**, Keele University and NICTA, 2004.

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. **Information and Software Technology (IST)**, v. 55, n. 12, p. 2049-2075, 2013.

KITCHENHAM, B.; BRERETON, P.; BUDGEN, D.; TURNER, M.; BAILEY, J.; LINKMAN, S. Systematic literature reviews in software engineering - A systematic literature review. **Information and Software Technology (IST)**, v. 51, n.1, p. 7-15; 2013.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. **Technical Report (EBSE) 2007-001**, Keele University and Durham University Joint Report, UK, 2007.

KITCHENHAM, B.; DYBA, T.; JØRGENSEN, M. Evidence-based software engineering. In: **Proceedings of 26th International Conference on Software Engineering (ICSE)**, Washington, USA, p. 273-281, 2004.

KITCHENHAM, B.; PRETORIUS, R.; BUDGEN, D.; BRERETON, O. P.; TURNER, M.; NIAZI, M.; Linkman S. Systematic literature reviews in software engineering - A tertiary, **Information and Software Technology (IST)**, v. 52, n.8 , p. 792-805, 2010.

LANDIS, J.; KOCH, G. The measurement of observer agreement for categorical data. **Biometrics (IBS)**, v. 33, n. 1, p. 159-174, 1977.

MACDONELL, S.; SHEPPERD, M.; KITCHENHAM, B.; MENDES, E. How Reliable Are Systematic Reviews in Empirical Software Engineering. **IEEE Transactions on Software Engineering (TOSEM)**. v.36, n. 5, p. 676-687, 2010.

MAFRA, S. N.; TRAVASSOS, G. H. estudos primários e Secundários Apoiando a Busca por Evidência em Engenharia de Software. **Technical Report RT-ES**, Programa de Engenharia de Sistemas e Computação. Universidade Federal do Rio de Janeiro, Rio de Janeiro, BR, 2006.

MARSHALL, C.; BRERETON, P. Systematic Review Toolbox: A Catalogue of Tools to Support Systematic Reviews. In: **19th International Conference on**

Evaluation and Assessment in Software Engineering (EASE). Nanjing, China: ACM, p. 1-6, 2015.

MARSHALL, C.; BRERETON, P.; KITCHENHAM, B. Tools to Support Systematic Reviews in Software Engineering: A Feature Analysis. In: **Proceedings of 18th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, London, United Kingdom: ACM, p.1-10, 2014.

MARSHALL, C.; BRERETON, P. Tools to Support Systematic Literature Reviews in Software Engineering: A Mapping Study. In: **Proceedings of 7th International Symposium on Empirical Software and Measurement (ESEM)**, Maryland, USA: ACM, p. 139-148, 2015.

MARSHALL, C.; BRERETON, P.; KITCHENHAM, B. Tools to Support Systematic Reviews in Software Engineering: A Cross-Domain Survey using Semi-structured Interviews. In: **Proceedings of 19th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, Nanjing, China: ACM, p. 1-6, 2015.

OCTAVIANO, F.; FELIZARDO, K.; MALDONADO, J., FABBRI, S. Semi-automatic selection of primary studies in systematic literature reviews: is it reasonable?.**Empirical Software Engineering (JESE)**. v20, n. 6, p. 1898-1917, 2015.

OCTAVIANO, F.; SILVA C.; FABBRI S. Using the SCAS strategy to perform the initial selection of studies in systematic reviews: an experimental study. In **Proceedings of 20th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, Limerick, Ireland: ACM, p 1-10, 2016.

PETERSEN, K.; FELDT R.; MUJTABA S.; MATTSSON, M. Systematic mapping studies in software engineering. In: **Proceedings of 12th international conference on Evaluation and Assessment in Software Engineering (EASE)**, Swindon, United Kingdom: ACM, p. 68-77. 2008.

PETERSEN K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. **Information and Software Technology (IST)**, v.64, n. c, p. 1-18, 2015.

SALTON, G.; ALLAN, J. Text retrieval using the Vector Processing Model. In: **Symposium on Document Analysis and Information Retrieval**, Information Science Research Institute, Las Vegas, USA, p. 9-22. 1994.

SCOPUS, **SCOPUS Digital Library**, 2017, Disponível em <<https://www.scopus.com/>>. Acesso em: 17 ago. 2017.

SEDGEWICK, R. **Algorithms in C++: Part 5 Graph Algorithms**, 3° edição, MA: Addison-Wesley, 2002.

SOUZA, R.; Lopes, C.; Bezerra, F.; de Souza, C. R. B. RAMANI: Uma Ferramenta de Apoio à Colaboração. In: **Proceedings of 13th Brazilian Symposium in Collaborative Systems (SBSC)**. Manaus, Brazil: ACM, p.144-147, 2013.

STAPLES, M; NIAZI, M. Experiences Using Systematic Review Guidelines. **Systems and Software (JSS)**. v.80; n. 9, p. 1425-1437, 2007.

WEBSTER, J.; WATSON, R. T. Analyzing the past to prepare for the future: Writing a literature review. **MIS Quarterly (MISQ)**, v. 26, n. 2, p. xiii-xxiii, 2002.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering In: **Proceedings of 18th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, London, United Kingdom: ACM, p.1-10, 2014.

WOHLIN, C. Second-generation systematic literature studies using snowballing. In: **Proceedings of 20th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, Limerick, Ireland: ACM, p.1-6, 2016.

WOHLIN, C.; RUNESON, P., NETO, P. A. M. S, ENGSTRÖMB, E., MACHADO, I. C.; DE ALMEIDA, E. S. On the reliability of mapping studies in software engineering. **Systems and Software (JSS)**. v. 86, n. 10 p. 2594–2610, 2013.

ZHANG, H.; BABAR, M. A. On Searching Relevant Studies in Software Engineering In: **Proceedings of 14th International Conference on Evaluation and Assessment in Software Engineering (EASE)**, p.111-120, 2010.

ZHANG, H.; BABAR, M. A.; TELL P. Identifying relevant studies in software engineering. **Information and Software Technology (IST)**, v. 53, n. 6, p.625-637, 2011.

APÊNDICE A

PROTOCOLO DO MAPEAMENTO SISTEMÁTICO

O Protocolo referente ao MS descrito na Seção 3.3 do Capítulo 3 é apresentado na Tabela 35.

Tabela 35. Protocolo do mapeamento sistemático

Title	O uso de Snowballing em estudos secundários: Um Mapeamento Sistemático
Researchers	Cleiton Silva; Sandra Fabbri;
Description	O Mapeamento deverá explorar a forma como a técnica Snowballing é aplicada em Revisão Sistemática (RS) e Mapeamento Sistemático(MS) em diferentes áreas.
Objectives	Identificar o Estado da Arte da técnica Snowballing como recurso de execução de buscas em Revisões Sistemáticas (RS) e Mapeamentos Sistemáticos (MS). Como a comunidade utiliza a técnica Snowballing durante a execução de RS ou MS
Main Question	
Intervention	Snowballing
Control	Retornar os artigos: 1. Guidelines for snowballing in systematic literature studies and a replication in software engineering 2. Systematic literature studies: database searches vs. backward snowballing
Population	Estudos que utilizem a técnica snowballing para o planejamento / execução de uma RS ou MS e descrevam como foi utilizado, ou estudos que tratem do processo de Snowballing aplicado a RS ou MS
Results	Caracterizar a técnica Snowballing aplicada a RS e MS
Application	Pesquisadores e Aplicadores que estejam utilizando/avaliando ou executando RS e MS que pretendam utilizar a técnica Snowballing
Keywords	bola de neve; bola de nieve; lista de referencia; reference list; reference tracking; snow balling; snowballing;
Source Selection	Indexar principais periódicos e publicações da área de computação ou medicina;

Criteria	
Studies Languages	Inglês; Português; Espanhol;
Source Search Methods	Serão utilizadas strings de busca equivalentes baseadas em uma string inicial e aplicadas nas máquinas de busca;
Source Engine	Web of Science; Science Direct; Scopus; Lilacs; Cinahl; PubMed; Cochrane;
Studies inclusion and exclusion criterias	(I) Apresenta "Guidelines" para a execução de RS ou MS usando Snowballing; (I) Estudo utiliza a técnica snowballing e descreve como foi utilizada; (E) Não Utiliza a técnica Snowballing; (E) Utiliza a técnica Snowballing em outro contexto (além de RS e MS); (E) Não é um estudo;
Studies types definition	Utilizar Qualquer tipo de estudo;
Initial studies selection	Os estudos deverão ser selecionados levando-se em consideração título, resumo e palavras-chave.
Studies quality evaluation	Aceitar qualquer estudo encontrado sobre o tema. Entretanto a qualidade metodológica deverá ser avaliada.
Information Extraction Fields	Idioma={Português,Inglês,Espanhol}; O Estudo Conduz=[N/A, RS, MS, Metanálise, Outro]; O Artigo Aplica a técnica=[NA ou Nenhuma Técnica, Systematic Search, Snowballing, Gray Literature, Hand Search, Expert Opinion, Research Group]; Área de conhecimento do estudo ={Computação, Áreas da Medicina e da Saúde, Outras Áreas}; Estudo Descreve detalhadamente como a técnica snowballing foi executada={N/A, Não, Sim}; O Estudo aplica =[NA, snowballing forward, snowballing backward]; Descrever (Se Houver) Ferramentas ou Apoio computacional utilizado para aplicar a técnica Snowballing; Base de Dados=[N/A, Scopus, ACM, IEEE, Web of Science, Science Direct, Springer Link, PubMed, Lilacs, Central, AISEL, Inspec, Compendex, AIS, MEDLINE, EMBASE, PsychINFO, Single, SPORTDiscus, AMED, Cochrane, PEDro, SMART, Cinahl,Australian Medical Index, Sociological Abstracts, Health

	and Safety Science, Global Health, Health Management Information Consortium, ERIC, PsychExtra, International Pharmaceutical, Australasian Digital Thesis Program, USA Virginia Henderson Library for nursing, APAIS Health, Proquest Health, Wiley Interscience, Science Citation Index Expanded, Social Care Online, PsychLIT, Social Science Citation Index (SSCI), British Nursing Index, ASSIA, HMIC, OpenSIGLE, Social Services Abstracts, World Bank E-Library, Swemed, Google Scholar]; O Artigo deve ser incluído na RS={Não, Sim}; Informações Adicionais;
Results Summarization	Ao final do levantamento bibliográfico, espera-se identificar o estado da arte da aplicação da técnica Snowballing em de RS e MS;

APÊNDICE B

PROTOCOLO DA REVISÃO SISTEMÁTICA

O Protocolo referente a RS descrita na Seção 3.4 do Capítulo 3 é apresentado na Tabela 36.

Tabela 36. Protocolo da revisão sistemática

Title	<i>Snowballing</i> aplicado em estudos secundários: Uma Revisão Sistemática
Researchers	Cleiton Silva; Sandra Fabbri;
Description	A revisão deverá explorar a forma como a técnica <i>Snowballing</i> é aplicada no processo de Revisão Sistemática (RS) e mapeamento Sistemático (MS)
Objectives	Identificar o Estado da Arte da técnica <i>Snowballing</i> como recurso de execução de busca em Revisões Sistemáticas (RS) e Mapeamentos Sistemáticos (MS).
Main Question	Como a comunidade propõe que seja utilizada a técnica <i>snowballing</i> para execução de busca em RS ou MS
Intervention	Snowballing
Control	Retornar os artigos: 1. Guidelines for snowballing in systematic literature studies and a replication in Software Engineering 2. Systematic literature studies: database searches vs. backward snowballing
Population	estudos primários que tratem do processo de Snowballing aplicado a RS ou MS
Results	Caracterizar a técnica <i>Snowballing</i> aplicada a RS e MS
Application	Pesquisadores e Aplicadores que estejam utilizando/avaliando ou executando RS e MS que pretendam utilizar a técnica <i>Snowballing</i>
Keywords	bola de neve; bola de nieve; lista de referencia; <i>reference list</i> ; <i>reference tracking</i> ; <i>snow balling</i> ; <i>Snowballing</i> ;
Source Selection Criteria	Possuir o "Full Text" do artigo disponível para <i>download</i> ; Indexar principais periódicos e publicações da Área;

Studies Languages	Inglês; Português; Espanhol;
Source Search Methods	Serão utilizadas strings de busca equivalentes baseadas em uma <i>String</i> inicial e aplicadas nas máquinas de busca;
Source Engine	Os estudos foram obtidos pelo mapeamento prévio.
Studies inclusion and exclusion criterias	(I) Apresenta " <i>Guidelines</i> " para a execução de RS ou MS usando <i>Snowballing</i> ; (I) Estudo Apresenta Suporte Computacional para execução de <i>Snowballing</i> ; (E) Não Utiliza a técnica <i>Snowballing</i> ; (E) Utiliza a técnica <i>Snowballing</i> em outro contexto (além de RS e MS); (E) Não possui o <i>FullText</i> Disponível e não aparenta tratar " <i>Guidelines</i> " para <i>Snowballing</i> ;
Studies types definition	Utilizar estudos primários de qualquer natureza.
Initial studies selection	Os estudos deverão ser selecionados levando-se em consideração título, resumo e palavras-chave;
Studies quality evaluation	Aceitar qualquer estudo encontrado sobre o tema desde que trate de uma ferramenta, ou descreva o processo de <i>Snowballing</i> explicitando como a proposta garante melhores resultados ao executar a técnica.
Information Extraction Fields	Idioma={Português,Inglês,Espanhol}; Grupo de Pesquisa; O estudo faz parte do " <i>Seed</i> " Inicial ={Não,Sim}; O estudo foi obtido por ={Mapeamento Prévio,Snowballing,Gray Literature,Research Group,Hand Search}; Área de conhecimento do estudo ={Computação, Áreas da Medicina e da Saúde,Outras Áreas}; Estudo Descreve detalhadamente como a técnica <i>Snowballing</i> foi (deve ser) executada={N/A,Não, Sim}; O Estudo aplica =[NA, <i>snowballing forward</i> , <i>snowballing backward</i>]; O Estudo Utiliza apoio Computacional para <i>Snowballing</i> ={N/A,Não,Sim}; Descrever (Se Houver) Ferramentas ou Apoio computacional utilizado para aplicar a técnica <i>Snowballing</i> ; Descrever (Se Houver) a metodologia utilizada para a execução da técnica <i>Snowballing</i> ; O estudo apresenta vantagens em utilizar o <i>Snowballing</i> ={N/A,Sim,Não}; Descrever (Se Houver) Vantagem de Utilizar a técnica <i>Snowballing</i> ; O estudo apresenta desvantagens em utilizar o <i>Snowballing</i> ={N/A,Sim,Não};

	Descrever (Se Houver) Desvantagem de Utilizar a técnica <i>Snowballing</i> ; Informações Adicionais;
Results Summarization	Ao final do levantamento bibliográfico, espera identificar o Estado da Arte da aplicação da técnica <i>Snowballing</i> em RS e MS;

APÊNDICE C

A FERRAMENTA StART

A Ferramenta StArt (State of the Art through Systematic Reviews) é uma ferramenta de suporte à condução de estudos secundários (Fabbri et. a., 2012). A ferramenta foi desenvolvida e é mantida pelo Laboratório de Pesquisas em Engenharia de Software (LaPES) da Universidade Federal de São Carlos (UFSCar).

A StArt tem por objetivo auxiliar o pesquisador a conduzir o levantamento bibliográfico de forma sistemática, oferecendo suporte para todas as etapas de estudos secundários, desde a construção de um protocolo seguindo o proposto por Kitchenham (2004), até a sumarização dos resultados.

Hernandes *et al.* (2012) fazem uma avaliação da ferramenta StArt, que mostra que a ferramenta auxilia positivamente na condução de RS. A Figura 16 exibe a interface da ferramenta StArt. Na região demarcada por (A), é possível selecionar as fases e etapas do processo de Revisão. Na região demarcada por (B) é apresentada uma interface que auxilia o pesquisador com relação às atividades requeridas em cada uma das etapas.

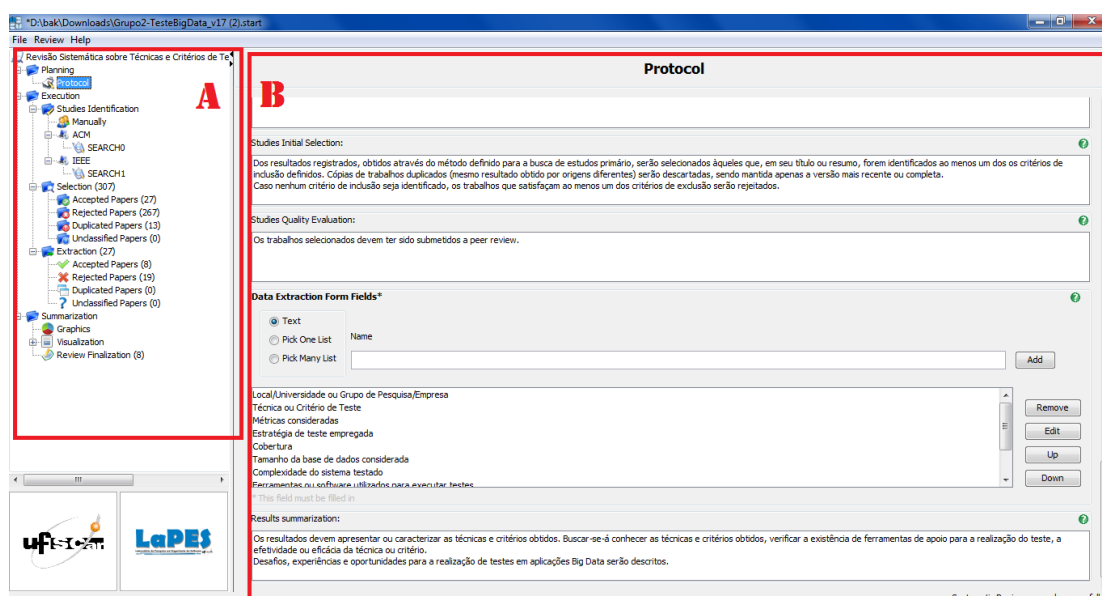


Figura 16 Plano de Trabalho da Ferramenta StArt

A ferramenta oferece suporte parcial a Busca com *Strings*, a qual deve ser executada em máquinas de busca e exportada para formatos de referências. A ferramenta permite a importação de arquivos de referências em diferentes formatos, tais como BibTex, RIS e Medline. Na Figura 17, é possível observar a interface que a ferramenta oferece para a catalogação de estudos. Na região demarcada por (C), a ferramenta armazena dados referentes à *string* utilizada na busca. Na região (D), a interface para a importação dos arquivos de referência obtidos das máquinas de busca, e, na região (E) a interface com os estudos catalogados, contendo

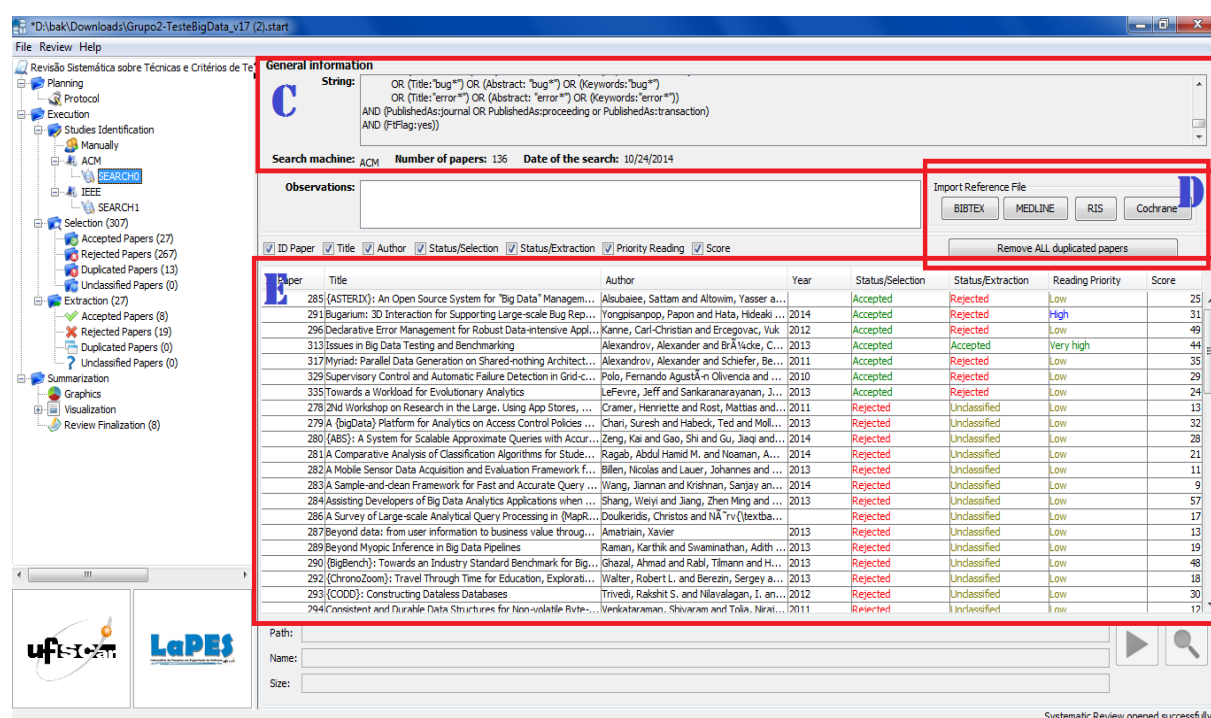


Figura 17 Catalogação de estudos na ferramenta StArt

informações básicas dos mesmos, tais como: título, autor, resumo, ano e palavras-chaves.

A ferramenta oferece recursos de visualização para auxiliar o pesquisador. As visualizações disponíveis são divididas em três grupos: grafo, árvore e grafo radial. As visualizações podem ser construídas com as informações dos metadados dos estudos localizados ou com as informações extraídas pelo pesquisador.

A interface para visualização dos dados é apresentada na Figura 18, na região (F), é possível filtrar os dados de acordo com os critérios definidos pelo pesquisador para cada referência. Em (G), os dados são agrupados de acordo com

informações disponíveis. Na região (H), é possível escolher o tipo de visualização presente na ferramenta e, em (I), é mostrada a visualização com os dados da pesquisa. Fabbri et al. (2012) apresenta alguns dos recursos de mineração de texto e identificação das referências das referências disponível na StArt.

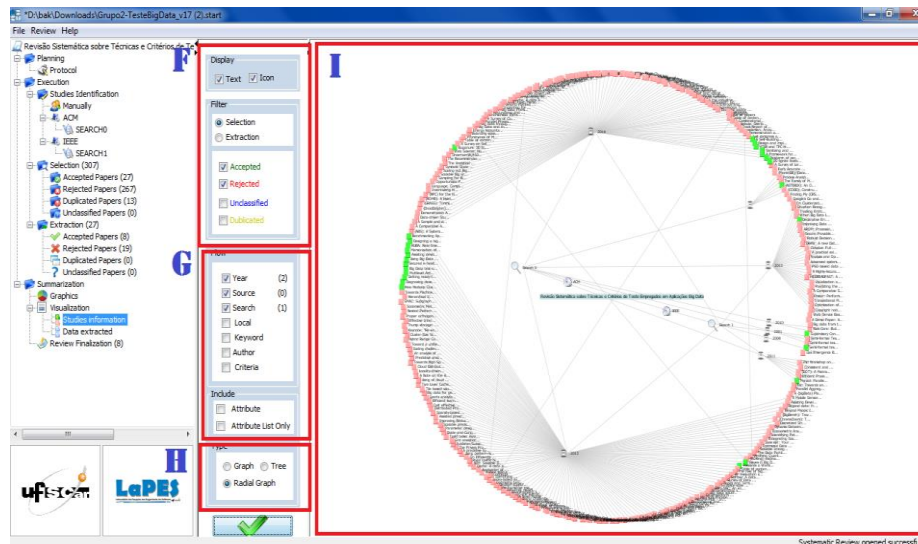


Figura 18. Interface para visualização de dados na StArt

A Figura 19 apresenta a interface com as informações detalhadas de cada estudo. Na região (J), o pesquisador pode obter dados do estudo, definir critérios de inclusão e exclusão, classificar e definir prioridades, extrair dados com o formulário de extração e identificar estudos similares dentro da pesquisa por meio de algoritmos de mineração de texto que a ferramenta utiliza.

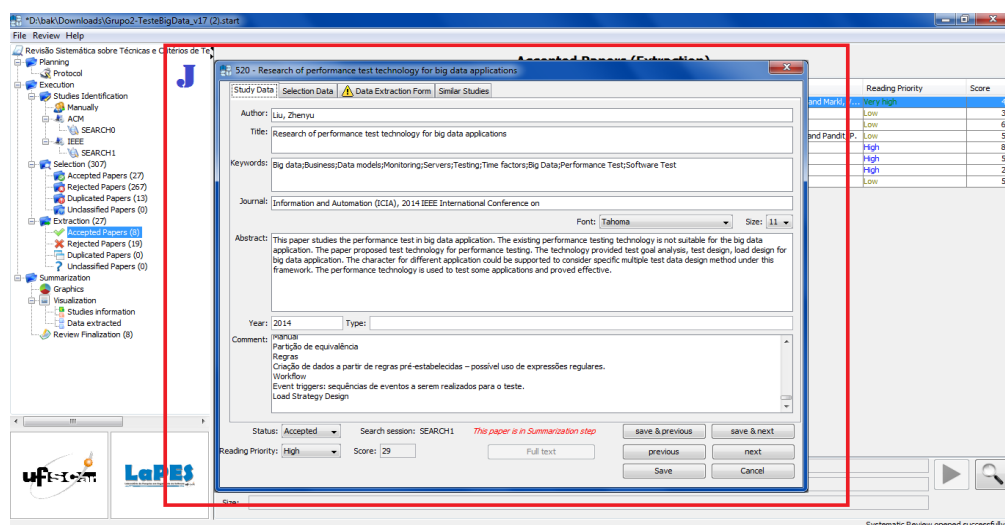


Figura 19. Interface para classificação de estudos na StArt

Existem diversos outros recursos de auxílio à condução de estudos secundários na ferramenta, entretanto, estes foram explicitados porque as modificações que serão implementadas na StArt para inclusão de recursos de *Snowballing* terá ligação, de alguma maneira, com os recursos apresentados.

APÊNDICE D

REFERÊNCIAS DOS ESTUDOS DE CASO

A Tabela 37 mostra todos os estudos aceitos por Hernandez, Belgamo e Fabbri (2013) no estudo exploratório apresentado na Seção 5.3 do Capítulo 5. Cada estudo recebeu uma identificação na primeira coluna (N°), a qual foi utilizada pelo autor para se referir aos mesmos neste trabalho.

Tabela 37. Conjunto de estudos aceitos por Hernandez, Belgamo e Fabbri (2013)

N°	Título	Autores
1	Early software product improvement with sequential inspection sessions: An empirical investigation of inspector capability and learning effects	Winkler, D., Thurnher, B., Biffi, S.
2	PBR vs. checklist: A replication in the N-fold inspection context	He, L., Carver, J.
3	Using observational pilot studies to test and improve lab packages	Mendonca, M.a , Cruzes, D.a , Dias, J.a , De Oliveira, M.C.F.b
4	A family of experiments to investigate the effects of groupware for software inspection	Biffi, S.a , Grunbacher, P.b , Halling, M.c
5	Improving the software inspection process	Stalhane, T., Awan, T.H.
6	Improvement of design specifications with inspection and testing	Winkler, D., Riedl, B., Biffi, S.
7	Investigating the impact of active guidance on design inspection	Winkler, D., Biffi, S., Thurnher, B.
8	ISPIS: A framework supporting software inspection processes	Kalinowski, M., Travassos, G.H.
9	A computational framework for supporting software inspections	Kalinowski, M., Travassos, G.H.
10	Improving software inspection process with patterns	Harjumaa, L.a , Tervonen, I.a , Vuorio, P.b
11	A controlled experiment for evaluating a metric-based reading technique for requirements inspection	Bernardez, B.a , Genero, M.b , DurÃ¡n, A.a , Toro, M.a
12	Eight maxims for software inspectors	Kelly, D., Shepard, T.
13	Assessing defect detection performance of interacting teams in object-oriented design inspection	Sabalaiuskaite, G.a c , Kusumoto, S.b , Inoue, K.b
14	Team-based fault content estimation in the software	Thelin, T.

	inspection process	
15	Distributed software inspections - An experiment with Adobe Acrobat	Harjumaa, L.
16	Tool support for geographically dispersed inspection teams	Lanubile, F.a b , Mallardo, T.a , Calefato, F.a
17	Empirical evaluations of usage-based reading and fault content estimation for software inspections	Thelin, T.
18	The development and evaluation of three diverse techniques for object-oriented code inspection	Dunsmor, A.a , Roper, M.b , Wood, M.b
19	Practical code inspection techniques for object-oriented systems: An experimental comparison	Dunsmore, A., Roper, M., Wood, M.
20	Computer-mediated group support, anonymity, and the software inspection process: An empirical investigation	Vitharana, P.a , Ramamurthy, K.b
21	Reducing inspection interval in large-scale software development	Perry, D.E.a c , Porter, A.b d , Wade, M.W.a e , Votta, L.G.a f , Perpich, J.e
22	On the independence of software inspectors	Miller, J.
23	Investigating the impact of reading techniques on the accuracy of different defect content estimation techniques	Freimut, B., Laitenberger, O., Biffli, S.
24	Estimating the number of remaining defects after inspection	Miller, J.
25	Perspective-based usability inspection: An empirical validation of efficacy	Zhang, Z.a b c , Basili, V.a b , Shneiderman, B.a d
26	Evaluating the usefulness and the ease of use of a web-based inspection data collection tool	Laitenberger, Oliver, Dreyer, Horst M.
27	Further Experiences with Scenarios and Checklists	Miller, J.a b c d , Wood, M.a d e , Roper, M.a b f
28	An experimental evaluation of an experience-based capture-recapture method in software code inspections	Runeson, P., Wohlin, C.
29	Communication and organization: An empirical study of discussion in inspection meetings	Seaman, C.B., Basili, V.R.
30	A comparison of tool-based and paper-based software inspection	Macdonald, F., Miller, J.
31	Understanding the Sources of Variation in Software Inspections	Porter, A.a , Siy, H.a , Mockus, A.b , Votta, L.b
32	A Replicated Experiment to Assess Requirements Inspection Techniques	Fusaro, P.a c d e , Lanubile, F.b c f g h , Visaggio, G.c g i j k l
33	Perspective-based reading of code documents at Robert Bosch GmbH	Laitenberger, O., DeBaud, J.-M.
34	A Process for asynchronous software inspection	Murphy, Paul, Miller, James
35	Comparing detection methods for software requirements inspections: a replicated experiment	Porter, Adam A., Votta Jr., Lawrence G., Basili, Victor R.
36	Can observational techniques help novices overcome the software inspection learning curve? An empirical investigation	Carver, Jeffrey C. and Shull, Forrest and Basili, Victor
37	An empirical study on design quality improvement from best-practice inspection and pair programming	Winkler, Dietmar and Biffli, Stefan
38	On the advantages of a systematic inspection for evaluating hyper media usability	De Angeli, A and Matera, M and Costabile, MF and Garzotto, F and Paolini, P
39	Experimentation with usage-based reading	Thelin, T and Erlansson, M and Host, M and Wohlin, C
40	Testing and inspecting reusable product line components: first empirical results	Denger, Christian and Kolb, Ronny
41	Detecting defects in object-oriented designs: using reading techniques to increase software quality	Travassos, Guilherme and Shull, Forrest and Fredericks, Michael and Basili, Victor R
42	A paradigm for metric based inspection process for	Nair, T. R. Gopalakrishnan and

	enhancing defect management	Suma, V.
43	Maximising the information gained from an experimental analysis of code inspection and static analysis for concurrent java components	Wojcicki, Margaret A and Strooper, Paul
44	Systematic object-oriented inspection: an empirical study	Dunsmore, Alastair and Roper, Marc and Wood, Murray
45	Evaluation of capture-recapture models for estimating the abundance of naturally-occurring defects	Walia, Gursimran Singh and Carver, Jeffrey C
46	Systematic inspection of information visualization systems	Ardito, Carmelo and Buono, Paolo and Costabile, Maria F and Lanzilotti, Rosa
47	Defects in automotive use cases	Torner, F. and Ivarsson, M. and Pettersson, F. and Ohman, P.
48	Systematic evaluation of e-learning systems: an experimental validation	Ardito, C. and Costabile, M. F and De Angeli, A. and Lanzilotti, R.
49	Improving software inspections with group process support	Tyran, Craig K and George, Joey F
50	An experiment to assess the cost-benefits of code inspections in large scale software development	Porter, A. and Siy, H. and Toman, C. A and Votta, L. G
51	A tool to support usability inspection	Ardito, Carmelo and Lanzilotti, Rosa and Buono, Paolo and Piccinno, Antonio
52	Tailoring a COTS Group Support System for Software Requirements Inspection	Halling, Michael and Grunbacher, Paul and Biff, Stefan
53	Metaphors of human thinking for usability inspection and design	Frokjaer, Erik and Hornbaek, Kasper
54	Task-directed software inspection technique: an experiment and case study	Kelly, Diane and Shepard, Terry
55	A catalogue of lightweight visualizations to support code smell inspection	Parnin, Chris and Gurg, Carsten and Nnadi, Ogechi
56	Validating the defect detection performance advantage of group designs for software reviews: report of a laboratory experiment using program code	Land, Lesley Pek Wee and Sauer, Chris and Jeffery, Ross
57	Further investigations into the development and evaluation of reading techniques for object-oriented code inspection	Dunsmore, Alastair and Roper, Marc and Wood, Murray
58	An experimental study of fault detection in user requirements documents	Schneider, G. Michael and Martin, Johnny and Tsai, W. T
59	Towards an inspection technique for use case models	Anda, Bente and Soberg, Dag I. K
60	A case study in the use of defect classification in inspections	Kelly, Diane and Shepard, Terry
61	An experiment to investigate interacting versus nominal groups in software inspection	Kelly, Diane and Shepard, Terry
62	Two psychology-based usability inspection techniques studied in a diary experiment	Hornbaek, Kasper and Frokjaer, Erik
63	Overcoming the {NAH} syndrome for inspection deployment	Jalote, Pankaj and Haragopal, M.
64	An empirical study of Web-based inspection meetings	Lanubile, F. and Mallardo, T.
65	Software product improvement with inspection. A large-scale experiment on the influence of inspection processes on defect detection in software requirements documents	Biff, S. and Halling, M.
66	Investigating the cost-effectiveness of reinspections in software development	Biff, S. and Freimut, B. and Laitenberger, O.
67	An empirical study of modifying the Fagan inspection process and the resulting main effects and interaction effects among defects found, effort required, rate of preparation and inspection, number of team members and product 1st pass quality	Eickelmann, N.S. and Ruffolo, F. and Baik, J. and Anant, A.

68	Evoluindo um Assistente de Apoio ao Processo de Inspeção através de Análise Quantitativa e Qualitativa	Fábio Santos, Marcos Gomes, Horácio A. B. F. Oliveira, Tayana Conte
69	The use of reading technique and visualization for program understanding	Porto, D.; Mendonca, M.; Fabbri, S.;
70	Inspeção dos Diagramas de Atividades na Especificação de requisitos	Mello, Rafael, Pereira, Wallace, Travassos, Guilherme
71	Automatic checking of quality best practices in software development documents	Dautovic, A.a and Plsch, R.a and Saft, M.b
72	A controlled experiment to assess the effectiveness of inspection meetings	Bianchi, A. and Lanubile, F. and Visaggio, G.
73	Evaluating distributed inspection through controlled experiments	De Lucia, A. and Fasano, F. and Scanniello, G. and Tortora, G.
74	Investigating the influence of inspector capability factors with four inspection techniques on inspection performance	Biffi, S and Halling, M
75	Using reading techniques to focus inspection performance	Halling, M and Biffi, S and Grechenig, T and Kohle, M
76	Investigating the effect of a second software inspection cycle -Cost-benefit data from a large-scale experiment on reinspection of a software requirements document	Biffi, S and Halling, H and Kohle, M
77	Evaluation of inspectors' defect estimation accuracy for a requirements document after individual inspection	Biffi, S and Grechenig, T and Kohle, M
78	Analysis of the impact of reading technique and inspector capability on individual inspection performance	Biffi, S
79	Anywhere, anytime code inspections: Using the web to remove inspection bottlenecks in large-scale software development	Perpich, JM and Perry, DE and Porter, AA and Votta, LG and Wade, MW