

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

Faculdade de Ciências e Tecnologia
Campus de Presidente Prudente

Programa de Pós-Graduação em Ciência da Computação

Protocolo de Revisão Sistemática

Uma Revisão Sistemática sobre Aprendizado Automatizado de
Modificações de Código-Fonte

Bolsista: Leandro Ungari Cayres

Orientador: Prof. Dr. Rogério Eduardo Garcia

Presidente Prudente
Fevereiro/2019

Leandro Ungari Cayres

Protocolo de Revisão Sistemática

Uma Revisão Sistemática sobre Aprendizado Automatizado de Modificações de Código-Fonte

Protocolo elaborado para a realização da Revisão Sistemática acerca de Aprendizado Automatizado de Modificações de Código-Fonte referente ao Projeto de Mestrado do aluno Leandro Ungari Cayres vinculado ao Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual Paulista.

Orientador: Prof. Dr. Rogério Eduardo Garcia

Presidente Prudente
Fevereiro/2019

Sumário

1	Revisão Sistemática de Literatura	1
1.1	Informações Gerais	1
1.2	Questões de Pesquisa	2
1.3	Identificação de Estudos	2
1.3.1	<i>String</i> de Busca	2
1.3.2	Fontes de Busca	3
1.3.3	Estratégia de Busca	3
1.4	Seleção e Avaliação de Estudos	3
1.4.1	Critérios de Avaliação	4
1.4.2	Estratégia de Seleção dos Estudos	5
1.4.3	Extração dos Dados	5
1.5	Síntese dos Dados e Apresentação	6
	Referências Bibliográficas	8

Revisão Sistemática de Literatura

Este capítulo apresenta todas as informações relevantes para a condução da Revisão Sistemática de Literatura conduzida neste projeto de pesquisa.

1.1 Informações Gerais

A presente revisão sistemática é intitulada de “Uma Revisão Sistemática sobre Aprendizado e Sugestão Automatizada de Modificações de Código-Fonte utilizando o Histórico de Versão”, a qual tem como intuito identificar as principais técnicas e abordagens utilizadas no processo de aprendizagem de modificações e como classificar tais melhorias/degradações de código-fonte, com base no impacto gerado em métricas de qualidade, através da análise histórico de versão em repositórios de software.

O grupo de pesquisadores envolvidos é composto pelos seguintes membros:

- **Nome:** Leandro Ungari Cayres
E-mail: leandro.ungari@unesp.br
Currículo Lattes: <http://lattes.cnpq.br/5996829502147029>
- **Nome:** Bruno Santos de Lima
E-mail: bruno.s.lima@unesp.br
Currículo Lattes: <http://lattes.cnpq.br/2119168921461476>
- **Nome:** Rogério Eduardo Garcia
E-mail: rogerio.garcia@unesp.br
Currículo Lattes: <http://lattes.cnpq.br/8031012573259361>

A condução dessa revisão tem como intuito formar uma base de conhecimento que colabore com a condução dos estudos especiais, assim como base para o restante do curso de Mestrado.

1.2 Questões de Pesquisa

De modo a identificar os pontos de investigação dentro da revisão sistemática, foram elaboradas as seguintes questões de pesquisa:

- QP1: Como modificações são detectadas entre versões de código-fonte?
- QP2: Como são aplicadas modificações de código-fonte?
- QP3: Como identificar padrões comuns entre modificações de código-fonte?
- QP4: Como modificações de código-fonte em geral podem ser classificadas entre melhorias e degradações?
- QP5: Como a sugestão de modificações é feita ao usuário?

A elaboração das questões de pesquisa foi realizada com base nos critérios PICO apresentados na Tabela 1.2.

População	Técnicas e abordagens que analisam modificações em repositórios de software.
Intervenção	A utilização do histórico do código-fonte na realização de melhorias do código-fonte.
Comparação	Não se aplica.
Resultados	Visão geral dos estudos quanto ao uso do histórico de modificações para melhoria do software, quanto aos métodos e tecnologias utilizadas de modo que permita a realização do propósito do estudo.

Tabela 1.1: Critérios PICO.

1.3 Identificação de Estudos

1.3.1 String de Busca

O primeiro passo para a condução da busca consiste na determinação de termos relevantes para a composição da *string* de busca. A Tabela 1.3.1 apresenta as palavras-chave definidas:

Palavras-chave	"code transformation", "code change", "code edit", "version", "improvement", "degradation", "learning", "code history", "code pattern".
----------------	---

Tabela 1.2: Palavras-chave da revisão sistemática.

A partir desse conjunto de palavras-chave foi definida a seguinte *string* de busca para a execução da busca:

("code"OR "source code"OR "code pattern") AND
("edit"OR "change"OR "transformation") AND
(("learning") OR ("history"OR "version"))

1.3.2 Fontes de Busca

As fontes de busca selecionadas para realizar a busca por estudos são conhecidas por serem as principais bases bibliográficas em Ciência da Computação (Nakagawa et al., 2017) e são fontes de buscas nas quais os pesquisadores envolvidos na elaboração da revisão sistemática têm acesso.

A Tabela 1.3.2 apresenta as bases bibliográficas utilizadas:

Fonte de Busca	Endereço Online	Tipo
IEEE Xplore	ieeexplore.ieee.org	Base Bibliográfica
ACM Digital Library	dl.acm.org	Híbrida
Enginnering Village	http://www.engineeringvillage.com	Motor de Busca
Science Direct	https://www.sciencedirect.com/	Base Bibliográfica

Tabela 1.3: Fontes de busca da revisão sistemática.

1.3.3 Estratégia de Busca

As buscas foram realizadas de forma automatizada das bases bibliográficas, em que a *string* de busca foi aplicada aos campos de título, resumo e palavras-chave. Vale ressaltar que foram conduzidas buscas-piloto com o intuito de aprimorar a *string* de busca.

Também foi aplicada a estratégia *Snowballing Avante* em cada um dos artigos resultantes da busca automatizada, visando identificar novos estudos primários que possam ser relevantes para a Revisão Sistemática da literatura.

Todos os estudos selecionados estavam delimitados ao período de 2012 a 2019. Adicionalmente, não foram realizadas consultas a especialistas, assim como buscas manuais.

1.4 Seleção e Avaliação de Estudos

O processo de seleção é dividido em duas etapas: Seleção inicial, na qual critérios de seleção são aplicados em todos os estudos identificados após a leitura de seu título e resumo, os estudos só serão excluídos se eles atenderem a pelo menos um critério de exclusão. Seleção final, segunda etapa, na qual os estudo selecionados na etapa anterior são analisados com maiores detalhes após sua leitura completa seguindo os critérios de seleção. Após isso, os estudos incluídos são avaliados com base nos critérios de qualidade (Nakagawa et al., 2017).

1.4.1 Critérios de Avaliação

A seguir estão listados os critérios de inclusão e exclusão, os quais serão utilizados no processo de seleção dos estudos obtidos da busca nos mecanismos de pesquisa.

- **Critérios de Inclusão:**

1. O estudo apresenta análise do histórico de versões do código-fonte no contexto de modificações de código-fonte ou análise de qualidade de código.
2. O estudo apresenta técnicas, abordagens relativas a detecção de modificações no código-fonte.
3. O estudo propõe ou relata algo relativo a aprendizado automatizado de modificações com análise direta do código-fonte.
4. O estudo relaciona-se ao aprendizado de padrões de modificação de código-fonte.
5. O estudo relaciona-se a mensuração da qualidade de código em modificações de código-fonte.
6. O estudo analisa o impacto de melhorias, defeitos na qualidade do código-fonte.
7. O estudo relata o uso de métricas para análise e classificação da qualidade do código-fonte.
8. O estudo apresenta alguma estratégia de sugestão de modificações.

- **Critérios de Exclusão:**

1. O estudo não possui resumo.
2. O estudo somente está publicado como resumo ou pôster.
3. O estudo não está escrito em inglês.
4. O estudo é uma versão mais antiga do outro estudo já considerado.
5. O estudo não é um estudo primário.
6. Não foi possível ter acesso ao estudo.
7. O estudo foi publicado antes de 2012.
8. O estudo não apresenta análise do código-fonte direta, mas sim utiliza outras abordagens.
9. O estudo envolve modificações de código-fonte muito específicas de um dado contexto.
10. O estudo não é relacionado a área de Engenharia de Software.
11. O estudo não é relacionado a modificações de código-fonte.

ID	Questão
Q1	O estudo é baseado em pesquisa ou somente relato de experiência?
Q2	Os objetivos estão claramente definidos?
Q3	O contexto em que o trabalho está inserido foi apresentado claramente?
Q4	A abordagem/estratégia desenvolvida está descrita de forma clara?
Q5	Os pesquisadores analisaram as vantagens/desvantagens/limitações da abordagem/estratégia utilizada?
Q6	Qual a qualidade das questões de pesquisa elaboradas?
Q7	O estudo utiliza dados reais?
Q8	A obtenção e análise dos resultados foi descrita de forma clara?
Q9	Foi analisada a influência dos próprios pesquisadores nos resultados?
Q10	Há justificativa e discussão dos resultados?
Q11	Foram encontradas contribuições relevantes do estudo?

Tabela 1.4: Questões de Avaliação de Qualidade – adaptado de [Marçal et al. \(2016\)](#).

12. O estudo é relacionado a modificações de código, mas não visa identificar, aprender ou analisar o impacto delas.
13. O estudo envolve modificações de código, porém é voltado para outras aplicações como visualização.

1.4.2 Estratégia de Seleção dos Estudos

A estratégia para seleção de estudos está representada na Figura 1.4.2. Todos os estudos resultantes da busca automatizada serão organizados visando remover estudos duplicados. Após a remoção de possíveis duplicações será executado em três etapas:

- Eliminação de Duplicados
- Seleção Inicial
- Seleção Final

Toda etapa de seleção deve ser documentada utilizando planilhas eletrônicas, sendo uma planilha para cada etapa de seleção.

1.4.3 Extração dos Dados

As questões elaboradas para a extração dos dados correspondem a quatro níveis para avaliação da qualidade: (1) Filtro inicial; (2) Rigor; (3) Credibilidade e (4) Relevância. Cada estudo foi avaliado em cada questão no critério de "Não aceitável", "Fracamente aceitável" e "Aceitável", cuja pontuação respectivamente é 0, 0.5 e 1.

A Tabela 1.4.3 apresentada as categorias de classificação dos estudos:

Após a classificação dos estudos, somente os estudos que obtiveram "Boa Qualidade" e "Muito Boa Qualidade" foram selecionados. Por fim, de modo a extrair os dados dos estudos selecionados foi utilizado o conjunto de questões apresentado na Tabela 1.4.3.

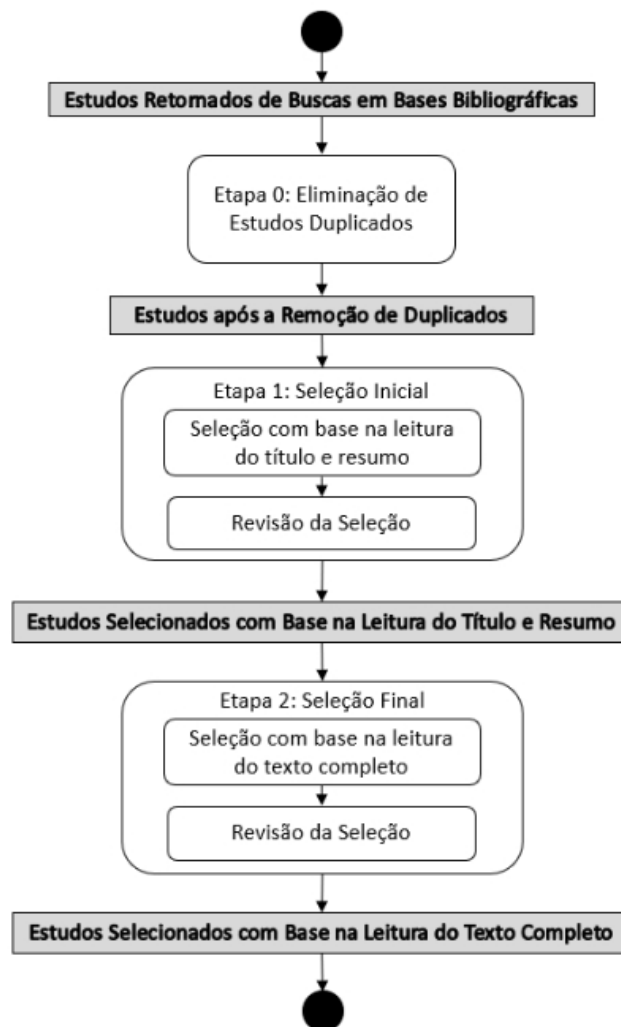


Figura 1.1: Sequência de atividades da etapa de seleção – adaptado de Nakagawa et al. (2017).

1.5 Síntese dos Dados e Apresentação

De modo a finalizar a atividade de seleção, é realizada a leitura completa dos estudos. Após essa leitura deve-se registrar os dados necessários que podem ser benéficos para responder as questões de pesquisa, que norteiam essa Revisão Sistemática (Petersen et al., 2008).

Para armazenar os dados extraídos e facilitar seu gerenciamento será utilizado uma planilha eletrônica utilizando a ferramenta WPS Spreadsheets ¹.

Quanto a atividade de sumarização dos resultados, tem-se como objetivo combinar dados extraídos dos estudos selecionados. Para sumarizar os dados foi utilizado o método qualitativo integrativo (Noblit et al., 1988, Nakagawa et al., 2017), em que os dados são sumarizados através do agrupamento de evidências.

Por fim, quanto a publicação, todos os resultados e dados obtidos serão descritos em artigo científico que será submetido para uma revista ou conferência que tenha na linha de pesquisa

¹<https://www.wps.com/spreadsheets>

1.5 Síntese dos Dados e Apresentação

Classificação	Pontuação
Muito Fraca Qualidade	0 - 3
Fraca Qualidade	3 - 6,5
Boa Qualidade	6,5 - 9
Muito Boa Qualidade	9 - 11

Tabela 1.5: Categorias de Classificação dos Estudos.

Inf 1	Identificador do Estudo
Inf 2	Título
Inf 3	Data de Extração
Inf 4	Autores
Inf 5	Ano
Inf 6	Visão Geral do Estudo
Inf 7	Contexto do Problema
Inf 8	Objectivo do Estudo
Inf 9	Estratégia da Abordagem/Técnica
Inf 10	Avaliação/Estudo de Caso/Experimento
Inf 11	Resultados
Inf 12	Avaliação dos Resultados
Inf 13	Conclusões e Contribuições

Tabela 1.6: Questão para Extração dos Dados.

relacionada a Aprendizado de Modificações de Código-Fonte em Engenharia de Software.

Além disso, uma Revisão Bibliográfica será apresentada para uma banca ao fim da disciplina de Estudos Especiais do Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Universidade Estadual Paulista.

Referências Bibliográficas

MARÇAL, I.; GARCIA, R. E.; ELER, D. M.; JUNIOR, C. O.; CORREIA, R. C. Techniques for the identification of crosscutting concerns: A systematic literature review. In: *Information Technology: New Generations*, Springer, p. 569–579, 2016.

NAKAGAWA, E. Y.; SCANNAVINO, K. R. F.; FABBRI, S. C. P. F.; FERRARI, F. C. *Revisão sistemática da literatura em engenharia de software: Teoria e prática*. Elsevier Brasil, 2017.

NOBLIT, G. W.; HARE, R. D.; HARE, R. *Meta-ethnography: Synthesizing qualitative studies*, v. 11. sage, 1988.

PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. Systematic mapping studies in software engineering. In: *EASE*, 2008, p. 68–77.