

MIRELLA DOS ANJOS SENA

**USANDO MAPEAMENTOS SISTEMÁTICOS PARA A ANÁLISE
DE GRUPOS DE PESQUISA EM ENGENHARIA DE SOFTWARE**

Dissertação apresentada ao Programa Multiinstitucional de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia e Universidade Estadual de Feira de Santana, como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientadores: Prof. Dr. Manoel Gomes de Mendonça Neto

Prof. Msc. Thiago Souto Mendes

Salvador

2017

Sena, Mirella dos Anjos.

Usando mapeamento sistemático para análise de grupos de pesquisa em engenharia de software / Mirella dos Anjos Sena. – Salvador, 2017.

149f. : il.

Orientador: Prof. Dr. Manoel Gomes de Mendonça Neto.

Dissertação (mestrado) – Universidade Federal da Bahia, Instituto de Matemática, Mestrado Multiinstitucional em Ciência da Computação, 2017.

Referências bibliográficas.

1. Engenharia de Software Experimental. 2. Mapeamento Sistemático.

I. Mendonça Neto, Manoel Gomes de. II. Universidade Federal da Bahia, Instituto de Matemática. III. Título.

CDU – 004.41




Universidade Federal da Bahia
Universidade Estadual de Feira de Santana



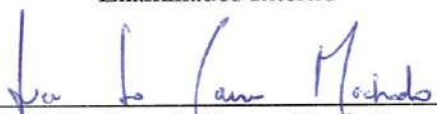
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

ATA DA 48ª DEFESA DE DISSERTAÇÃO DO MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

Aos vinte e sete dias do mês de abril de dois mil e dezessete, às 16h, na Sala de Videoconferência da Superintendência de Tecnologia da Informação da Universidade Federal da Bahia, foi instalada a sessão pública para julgamento da dissertação intitulada: “Usando Mapeamentos Sistemáticos para a Análise de Grupos de Pesquisa em Engenharia de Software”, elaborada pela mestranda Mirella dos Anjos Sena do Programa Multi-institucional de Mestrado em Ciência da Computação, matriculada no Programa sob número 213115416. Após a abertura da sessão, o Prof. Dr. Manoel Gomes de Mendonça Neto (UFBA), presidente da banca julgadora deu prosseguimento aos trabalhos, apresentando os demais examinadores da dissertação, os professores doutores: Prof. Dr. Ivan do Carmo Machado (UFBA) e a Profª. Dra. Sandra Camargo Pinto Ferraz Fabbri (UFSCAR). Após as apresentações, a palavra foi passada à autora da dissertação, Mirella dos Anjos Sena, que expôs seu trabalho dentro do tempo previsto de 50 minutos. Procedeu-se então à arguição da aluna pela banca e à leitura dos respectivos pareceres dos integrantes da banca. A aluna respondeu as questões levantadas pela banca de forma satisfatória. Ao término da arguição, a banca sugeriu algumas correções no trabalho e o presidente da banca abriu a palavra à platéia para comentários adicionais. Ao final, a banca reunida em separado, resolveu aprovar a dissertação da mestranda, concedendo-lhe o prazo de 60(sessenta) dias para realizar as correções sugeridas. Nada mais havendo a tratar, foi encerrada a sessão e lavrada a presente ata que é assinada por quem de direito. Salvador, 27 de abril de 2017.



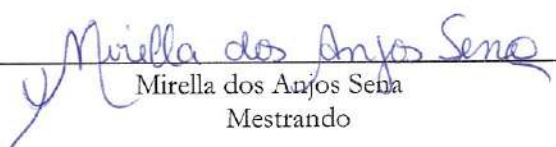
Prof. Dr. Manoel Gomes de Mendonça Neto
Examinador Interno



Prof. Dr. Ivan do Carmo Machado
Examinador Interno



Prof. Dr. Sandra Camargo P. Ferraz Fabbri
Examinadora Externa



Mirella dos Anjos Sena
Mestrando

RESUMO

As publicações realizadas por grupos, laboratórios ou comunidades de pesquisa (artigos, dissertações, relatórios técnicos) vêm aumentando na área de Engenharia de Software (ES). As pesquisas desenvolvidas por esses grupos são críticas para a sua consolidação e referência na comunidade científica. Neste contexto, o mapeamento sistemático é um método experimental que apresenta um processo de análise de dados coletados de estudos primários. Os resultados desse processo são apresentados por meio de mapas visuais, o que facilita o entendimento dos pesquisadores sobre uma área de pesquisa. Esse método, utilizado em estudos de ES, também pode ser utilizado para analisar publicações de um laboratório, grupo ou comunidade de pesquisa.

Este trabalho apresenta o método MASG – Método para Análise Sistemática de Grupos de Pesquisa – que possibilita caracterizar grupos, laboratórios ou comunidades de pesquisa, com o propósito de auxiliar seus líderes e pesquisadores em tomadas de decisão quanto ao direcionamento das novas pesquisas a serem realizadas, bem como identificar estratégias de pesquisa utilizadas pelo grupo.

Para avaliação do método foi realizado um estudo de caso com o Laboratório de Engenharia de Software (LES) da Universidade Federal da Bahia (UFBA) com foco na área temática de Mineração de Repositório de Software (MSR). Como resultado foi possível observar que o método proposto permite caracterizar pesquisas de um grupo, identificar o seu perfil e estratégias de pesquisa. Ele também permite que os líderes e pesquisadores identifiquem novas oportunidades de pesquisa a serem exploradas pelo grupo por meio da análise e comparação de gráficos e mapas visuais, bem como ganhem entendimento quanto à posição do grupo em relação às pesquisas que estão sendo realizadas em uma área de interesse. Também foram realizadas comparações individuais dos pesquisadores. Este processo caracterizou e posicionou os pesquisadores em relação ao grupo e à área analisada, e forneceu aos mesmos subsídios para tomadas de decisões quanto às suas pesquisas futuras.

Palavras chave: Mapeamento Sistemático, Análise Sistemática, Grupos de Pesquisa, Estudos Empíricos, Mineração de Repositório de Software.

ABSTRACT

The publications made by groups, laboratories or research communities (articles, dissertations, technical reports) have been increasing in the area of Software Engineering (ES). The researchers developed by these groups are critical for their consolidation and reference in the scientific community. In this context, systematic mapping is an experimental method that presents a process of analysis of data collected from primary studies. The results of this process are presented through visual maps, which facilitates the researchers' understanding of a research area. This method, used in ES studies, can also be used to analyze publications from a laboratory, group or research community.

This work presents the MASG method - Method for Systematic Analysis of Research Groups - that allows to characterize groups, laboratories or research communities, with the purpose of assisting its leaders and researchers in decision making as to the direction of the new researches to be carried out, as well as identify research strategies used by the group.

To evaluate the method, a case study was carried out with the Software Engineering Laboratory (LES) of the Federal University of Bahia (UFBA), focusing on Software Repository Mining (MSR). As a result it was possible to observe that the proposed method allows to characterize researches of a group, to identify its profile and research strategies. It also allows leaders and researchers to identify new research opportunities to be explored by the group through the analysis and comparison of graphs and visual maps, as well as to gain an understanding of the position of the group in relation to the research being carried out in an area of interest. Individual comparisons of the researchers were also performed. This process characterized and positioned the researchers in relation to the group and area analyzed, and provided the same subsidies for decision making regarding their future research.

Keywords: Systematic Mapping, Systematic Analysis, Research Groups, Empirical Studies, Software Repository Mining.

Dedico este trabalho a minha família, amigos e orientadores pelo apoio, força, incentivo e companheirismo.

AGRADECIMENTOS

Compartilhar o conhecimento é algo que enriquece a quem o recebe e mais ainda a quem o transmite. Por esse motivo agradeço a meu orientador o Prof. Dr. Manoel Gomes de Mendonça Neto por ter me guiado com a sua inteligência, paciência e conhecimento. Ele é uma pessoa indescritível. Agradeço a todo o momento por tê-lo como meu orientador e pela oportunidade que ele me deu de dar seguimento no mestrado. Sem o seu apoio esse trabalho não seria realidade.

Ao Prof. Msc. Thiago Souto Mendes meu “coorientador” e sempre amigo que me mostrou que todos nós temos pontos fortes e que os pontos fracos podem ser melhorados. Quem tiver a oportunidade de tê-lo como orientador vai entender o que sinto nesse momento. Muito obrigada.

A meu esposo pela paciência e incentivo de cada dia. Pelos dias perdidos e pelas viagens não realizadas. Acredito que o ganho foi muito maior em conhecimento durante essa prova que passamos.

A minha família, amigos em especial a Glaucya Boechat, Ivan Machado e Mário Farias e aos pesquisadores do LES (Laboratórios de Engenharia de Software da UFBA) pela atenção e disponibilidade de tempo cedido para a minha pesquisa.

A Deus meu único e principal influenciador e Senhor, pois sem Ele não conseguiria chegar até aqui.

Aos membros da banca pela atenção dada ao aceitar o convite, leitura e comentários para melhoria deste trabalho.

Agradeço ainda a todos aqueles que compartilharam comigo momentos de dedicação e, sobretudo, alegria ao longo destes anos.

É muito melhor lançar-se em busca de conquistas grandiosas, mesmo expondo-se o fracasso, do que alinhar-se com os pobres de espírito, que nem gozam muito nem sofrem muito, porque vivem numa penumbra cinzenta, onde não conhecem nem vitória, nem derrota.

Theodore Roosevelt

SUMÁRIO

1	INTRODUÇÃO.....	1
1.1	OBJETIVOS	3
1.1.1	Objetivo Geral	3
1.1.2	Objetivos Específicos	3
1.2	TRABALHO REALIZADO E RESULTADOS	4
1.3	ORGANIZAÇÃO DO TRABALHO.....	5
2	REFERENCIAL TEÓRICO	6
2.1	ESTUDOS SECUNDÁRIOS.....	6
2.2	REVISÃO SISTEMÁTICA DA LITERATURA.....	8
2.2.1	Planejamento da Revisão	10
2.2.2	Execução da Revisão	11
2.2.3	Análise da Revisão	13
2.3	MAPEAMENTO SISTEMÁTICO.....	13
2.3.1	Processo do Mapeamento Sistemático.....	14
2.3.2	Definição das questões de pesquisa.....	16
2.3.3	Condução da pesquisa.....	16
2.3.4	Triagem dos artigos	17
2.3.5	Resumo das palavras-chave.....	17
2.3.6	Extração dos dados e o processo de mapeamento	18
2.4	DIFERENÇAS ENTRE MAPEAMENTO SISTEMÁTICO E REVISÃO SISTEMÁTICA.....	19
2.5	TIPOS DE PESQUISA	21
2.5.1	Técnicas de Coleta de Dados Qualitativos	22
2.5.2	Entrevistas.....	23
2.5.2.1	Tipos de Entrevistas	24
2.6	TÉCNICAS PARA ANÁLISE DE DADOS QUALITATIVOS.....	25
2.6.1	Síntese Temática.....	26
2.7	CONSIDERAÇÕES FINAIS.....	31

3 O MASG.....	32
3.1 DESCRIÇÃO DO MÉTODO	32
3.1.1 Atividade 1 – Identificar a fonte de dados	35
3.1.2 Atividade 2 – Identificar os metadados e publicações do grupo.....	36
3.1.3 Atividade 3 – Selecionar as publicações do grupo.....	40
3.1.4 Atividade 4 – Extrair os dados das publicações do grupo.....	42
3.1.5 Atividade 5 – Analisar a classificação das publicações do grupo.....	44
3.1.6 Atividade 6 – Analisar e comparar os mapas	46
3.1.7 Atividade 7 – Entrevistar os pesquisadores do grupo	49
3.2 CONSIDERAÇÕES FINAIS	50
4 ARTEFATOS UTILIZADOS NO MÉTODO	51
4.1 MAPEAMENTO SISTEMÁTICO DA ÁREA.....	51
4.2 ROTEIRO DE ENTREVISTAS	53
4.3 TERMO DE CONSENTIMENTO.....	55
4.4 TABELA DE REFERÊNCIA PARA EXTRAÇÃO DOS DADOS DAS PUBLICAÇÕES	56
4.5 CONSIDERAÇÕES FINAIS	57
5 ESTUDO DE CASO.....	58
5.1 ANÁLISES DO GRUPO DE PESQUISA	59
5.1.1 Atividade 1 – Identificar a fonte de dados	59
5.1.2 Atividade 2 – Identificar os metadados e publicações do grupo.....	60
5.1.3 Atividade 3 – Selecionar as publicações do grupo.....	69
5.1.4 Atividade 4 – Extrair os dados das publicações do grupo.....	71
5.1.5 Atividade 5 – Analisar a classificação das publicações do grupo.....	71
5.1.6 Atividade 6 – Analisar e comparar os mapas	83
5.1.7 Atividade 7 – Entrevistar os pesquisadores do grupo	93
5.1.7.1 Dados quantitativos das entrevistas	94
5.1.7.2 Dados qualitativos das entrevistas.....	99
5.1.8 Extração dos dados.....	99
5.1.9 Codificação de Dados.....	102
5.1.10 Tradução de códigos em temas	104

5.1.11 Criação de um modelo de alto nível	105
5.2 CONSIDERAÇÕES FINAIS.....	107
6 CONCLUSÃO.....	109
6.1 DISCUSSÃO DOS RESULTADOS	109
6.2 RESULTADOS E CONTRIBUIÇÕES DESTE TRABALHO	110
6.3 LIMITAÇÕES	111
6.4 AMEAÇAS À VALIDADE	112
6.5 TRABALHOS FUTUROS	114
REFERÊNCIAS	115
APÊNDICE A	122
APÊNDICE B.....	123
ANEXO 1.....	125

LISTA DE FIGURAS

Figura 1. Etapas de uma revisão sistemática (Kitchenham <i>et al.</i> , 2007)	10
Figura 2. Etapas iniciais do mapeamento sistemático.....	15
Figura 3. Processo do mapeamento sistemático (Pertesen <i>et al.</i> , 2008)	15
Figura 4. Plano de classificação (Petersen <i>et al.</i> , 2008)	18
Figura 5. Método de síntese temática (adaptado de Creswell, 2007; Cruzes e Dyba, 2011a)	27
Figura 6. Método para Análise Sistemática de Grupos de Pesquisa (MASG)	34
Figura 7. Exemplo de gráfico para mostrar as publicações por pesquisador	38
Figura 8. Exemplo de gráfico para mostrar a quantidade de publicações por tipo	38
Figura 9. Exemplo de grafo de colaboração para mostrar a coautoria de publicações de pares de autores	39
Figura 10. Processo de seleção das publicações (adaptado de Neto <i>et al.</i> , 2011)	41
Figura 11. Exemplo de gráfico de bolhas para representar combinações de facetas	45
Figura 12. Exemplo de gráfico de colunas para comparar dados numéricos	47
Figura 13. Exemplo de sobreposição de gráficos de bolhas para visualizar diferenças entre resultados	48
Figura 14. Grupos que compõe o LES	60
Figura 15. Quantidade de produções do LES identificadas com o ScriptLattes em 19/07/2016	62
Figura 16. Quantidade e tipos de publicações analisadas	65
Figura 17. Quantidade de publicações do LES por tipo	66
Figura 18. Quantidade de publicações por pesquisador do LES.....	67
Figura 19. Grafo de colaboração dos pesquisadores do LES com peso	68
Figura 20. Triagem das publicações do LES	70
Figura 21. Proporção das publicações do LES entre 2010 e 2015 na área de MSR e das demais áreas de conhecimento	70
Figura 22. Publicações do LES por pesquisador na área de MSR	72
Figura 23. Publicações do LES por grupo de pesquisa na área de MSR	72

Figura 24. Grafo de colaboração com peso dos grupos do LES na área de MSR	73
Figura 25. Grafo de colaboração dos pesquisadores do LES que tiveram publicações na área de MSR.....	73
Figura 26. Grafo de colaboração dos pesquisadores que tiveram publicações conjuntas com os pesquisadores do LES na área de MSR	74
Figura 27. Quantidade de publicações do LES na área de MSR por pesquisador (mínimo 3) (autores/coautores).....	75
Figura 28. Propósitos utilizados pelas publicações do LES na área de MSR	76
Figura 29. Focos utilizados pelas publicações do LES na área de MSR	77
Figura 30. Objetivos de análise utilizados pelas publicações do LES na área de MSR	78
Figura 31. Quantidade de publicações do LES na área de MSR por tipo de fonte de dados	79
Figura 32. Gráficos de bolhas (foco x objeto de análise; foco x propósito) da área de MSR do LES	80
Figura 33. Fontes de dados utilizadas pelas publicações do LES na área de MSR por ano..	81
Figura 34. Métodos de avaliação utilizados pelas publicações do LES na área de MSR	82
Figura 35. Gráfico de bolhas (propósito x método de avaliação) do LES na área de MSR .	83
Figura 36. Comparação dos propósitos utilizados pelas publicações (MSR x LES)	84
Figura 37. Comparação dos focos utilizados pelas publicações (MSR x LES)	85
Figura 38. Comparação dos objetos de análise utilizados pelas publicações (MSR x LES).	86
Figura 39. Sobreposição dos resultados dos gráficos de bolhas (propósito x método de avaliação) (MSR x LES).....	87
Figura 40. Comparação das fontes de dados utilizadas pelas publicações (MSR x LES).....	88
Figura 41. Comparação das publicações que utilizam mais de uma fonte de dados (MSR x LES).....	88
Figura 42. Quantidade de publicações por tipo de fonte de dados e ano (MSR x LES)	89
Figura 43. Comparação dos métodos de avaliação utilizados nas publicações (MSR x LES)	90
Figura 44. Sobreposição de resultados dos gráficos de bolhas (propósito x foco; foco x objeto de análise) (MSR x LES)	92
Figura 45. Quantidade de pesquisadores entrevistados que tem a área de MSR como área principal ou secundária de pesquisa	95

Figura 46. Mapas mais citados nas entrevistas	97
Figura 47. Objetivos de pesquisa dos pesquisadores do LES	98
Figura 48. Exemplo de marcações feitas nos textos das entrevistas	100
Figura 49. Planilha de extração – aba Dados Gerais.....	101
Figura 50. Planilha de extração – aba Dados das Estratégias	102
Figura 51. Planilha de extração – aba Síntese Temática	103
Figura 52. Nível de interpretação na síntese temática (Cruzes e Dyba, 2011a)	104
Figura 53. Mapa Temático para identificar e classificar estratégias de pesquisa	106

LISTA DE QUADROS

Quadro 1. Refinamento do Método de Síntese Temática (Prates, 2016)	28
Quadro 2. Metadados do grupo	36
Quadro 3. Dados das publicações do grupo	40
Quadro 4. Exemplo resumido de tabela de classificação	44
Quadro 5 Roteiro de entrevistas	53
Quadro 6. Modelo do termo de consentimento para entrevistas	55
Quadro 7. Tabela de referência para extração dos dados das publicações	56
Quadro 8. Linhas de pesquisa por grupo do LES	61
Quadro 9. Tabela dos dados extraídos das entrevistas (adaptado de Prates, 2016).....	103

LISTA DE ABREVIATURAS E SIGLAS

BTS	<i>Bug Tracking System</i>
ITS	<i>Issues Tracking System</i>
LES	Laboratório de Engenharia de Software
MASG	Método para Análise Sistemática de Grupos
MS	Mapeamento Sistemático
MSR	Mineração de Repositório de Software
QE	Questão Específica
QG	Questão Geral
RS	Revisão Sistemática da Literatura
SoftVis	<i>Software Visualization</i> ou Visualização de Software

Neste capítulo é apresentada a introdução que aborda juntamente a motivação e o problema identificado para a realização desta pesquisa. Em seguida são apresentados o objetivo geral e os objetivos específicos, a metodologia e por último descrito como esta dissertação está organizada.

1 INTRODUÇÃO

Abordagens baseadas em evidências são bastante exploradas em decorrência dos seus resultados positivos nas áreas da medicina e ciências sociais. Estas abordagens buscam a síntese sistemática de evidências para contribuir de forma significativa e positiva para o avanço do estado-da-arte de uma área de pesquisa. Em Engenharia de Software (ES), a introdução do conceito de abordagens baseadas em evidências foi capitaneada por Kitchenham *et al.* (2004), em um conceito batizado como Engenharia de Software Baseada em Evidências (ESBE).

Este conceito utiliza-se de estudos experimentais, classificados como estudos primários (experimentos controlados, pesquisas de opinião e estudo de caso) e secundários (mapeamentos sistemáticos, revisões sistemáticas e meta-etnografia), para apoiar a análise de validade, impacto e aplicabilidade das evidências, por meio de sua avaliação crítica (Kitchenham *et al.*, 2004).

Estudos experimentais provêm um mecanismo de construção ampla de conhecimento, e podem ser utilizados por pesquisadores, bem como profissionais da indústria, para o apoio à tomada de decisões no desenvolvimento de soluções (Almeida *et al.*, 2011).

Nos últimos anos, o interesse pelo desenvolvimento de estudos secundários na área de ES tem crescido devido à sua capacidade de apresentar, por meio das análises de estudos primários, síntese de evidências que contribuem para melhorar o conhecimento dos pesquisadores e profissionais, bem como evidenciar lacunas de

pesquisas existentes em áreas de conhecimento (Kitchenham *et al.*, 2009; Kitchenham *et al.*, 2010; Almeida *et al.*, 2011; Da Silva *et al.*, 2011).

Existem métodos sistemáticos de síntese de evidência empírica que têm sido amplamente aplicados na área de ES. Dentre eles, destacam-se os mapeamentos sistemáticos, que apresentam uma visão geral de uma área ou objeto de estudo desejado. Com eles, evidências são apresentadas por meio de um resumo visual ou mapas, classificados e categorizados por resultados e dirigidos por meio de questões investigativas de pesquisa (Kitchenham *et al.*, 2007).

Em um Mapeamento Sistemático (MS) são analisados e mapeados estudos primários, disponíveis em bases de dados, referentes a uma área de conhecimento específica. Eles têm como foco conhecer um tópico de forma ampla, identificando quais lacunas estão presentes na área. Esses tipos de estudos são geralmente utilizados quando pesquisadores querem chegar a um entendimento comum do estado da arte de uma área de pesquisa, o que permite verificá-la em um determinado momento na busca de evidências.

Além disso, os mapeamentos sistemáticos também podem ser utilizados para caracterizar publicações de um laboratório, grupo ou comunidade de pesquisa (para fins de simplificar o entendimento, daqui em diante o texto mencionará somente o termo “grupos de pesquisa” suprimindo os demais), diferenciando-se dos mapeamentos sistemáticos, por exemplo, quanto à abrangência e aos trabalhos utilizados para análise, que pode incluir, neste segundo caso, todos os tipos de publicações desenvolvidas por seus pesquisadores (ex.: artigos, dissertações, teses, e relatórios técnicos).

Líderes de grupos de pesquisa podem enfrentar dificuldades para controlar a abrangência, posicionamento e foco das áreas de investigação científica sob sua coordenação, visto que grupos são cada vez mais heterogêneos em sua formação e os domínios de investigação científica crescem em ritmo geométrico (NSF, 2016).

Compreender o posicionamento de um grupo de pesquisa em uma determinada área de interesse pode apoiar a tomada de decisões dos líderes e ser utilizado como uma ferramenta de planejamento e gestão, em face à evolução de uma área de pesquisa.

Diante desse contexto, este trabalho propõe o MASG, um Método para Análise Sistemática de Grupos de Pesquisa, que utiliza dados de um MS pré-existente

de uma área de interesse a ser analisada, um conjunto de artefatos produzidos para o método, dados e publicações de um grupo de pesquisa. Este método possibilita a caracterização de um grupo de pesquisa na área de interesse, bem como a identificação das estratégias de pesquisa que vêm sendo utilizadas por ele. Essas informações são importantes no apoio aos líderes de grupos de pesquisa e de seus pesquisadores em tomadas de decisão quanto à posição e direção das novas pesquisas a serem realizadas pelo grupo.

1.1 OBJETIVOS

Nesta seção são apresentados o objetivo geral e os objetivos específicos desta pesquisa.

1.1.1 Objetivo Geral

O objetivo principal deste trabalho é propor um método para análise sistemática de grupos de pesquisa e de seus pesquisadores de forma individual para caracterizá-los em relação a uma determinada área de interesse, identificando no processo suas estratégias e atividades de pesquisa.

1.1.2 Objetivos Específicos

- Entender como informações de mapeamentos sistemáticos de uma área podem ser utilizadas para analisar publicações de um grupo de pesquisa;

- Definir um método de comparação entre os mapas de uma área e os mapas de um grupo de pesquisa;
- Definir como as comparações dos mapas podem ser usadas para caracterizar um grupo;
- Definir como as comparações dos mapas e outras informações coletadas podem ser usadas em entrevistas semiestruturadas para identificar as estratégias de pesquisas utilizadas por um grupo;
- Avaliar o processo acima definido em uma área de interesse específica, neste caso, Mineração de Repositório de Software (MSR), aplicando-o aos pesquisadores do Laboratório de Engenharia de Software (LES) da Universidade Federal da Bahia (UFBA).
- Criar um mapa temático para auxiliar na identificação e classificação das estratégias de pesquisa utilizadas por pelo grupo de pesquisa.

1.2 TRABALHO REALIZADO E RESULTADOS

A fim de aplicar o método proposto, foi realizado um estudo de caso com o LES da UFBA onde foram analisadas suas publicações relacionadas à área de MSR dos anos de 2010 a 2015. Os resultados indicam que com a utilização do método proposto as comparações feitas entre os mapas do LES e dos mapas produzidos em um MS da área de MSR apresentou uma visão de como o LES estava em relação a esta área e também de como os seus pesquisadores contribuem para a área e para o LES. A importância da análise foi confirmada por meio de entrevistas realizadas com os pesquisadores durante o processo.

As entrevistas com os pesquisadores do laboratório também permitiram a identificação das estratégias de pesquisa do LES, o que possibilitou a definição de um mapa temático para a realização desta atividade. Esse mapa temático pode ser utilizado para identificar e classificar estratégias de pesquisa adotadas por outros grupos na área estudada.

1.3 ORGANIZAÇÃO DO TRABALHO

Este Capítulo introduziu o tema de pesquisa deste trabalho – o desenvolvimento de um método para mapeamento de grupos de pesquisa. Ele descreveu a motivação para seleção do tema, o objetivo geral e os objetivos específicos do trabalho.

No Capítulo 2 é apresentada uma revisão bibliográfica do domínio, que contém os principais assuntos relacionados à pesquisa.

No Capítulo 3 é apresentado o MASG – Método para Análise Sistemática de Grupos de Pesquisa, principal contribuição deste trabalho.

No Capítulo 4 são apresentados os artefatos utilizados para a execução do método.

No Capítulo 5 é apresentado o estudo de caso utilizado para avaliar o método proposto.

Finalmente, no Capítulo 6 são discutidos os resultados e conclusão do trabalho. Nele também são apresentadas as contribuições e limitações do trabalho, lições aprendidas e as ideias para trabalhos futuros.

Os Apêndices A e B apresentam, respectivamente, a lista dos pesquisadores que foram analisados do LES e a lista de autores e coautores das publicações do LES relacionadas à área de MSR.

O Anexo 1 apresenta o MS da área de MSR de Farias *et al.* (2016) utilizado no estudo de caso para validação do MASG.

Neste capítulo é apresentada a revisão bibliográfica dos temas relacionados à pesquisa, com o objetivo de contextualizar os conceitos explorados durante o desenvolvimento desta dissertação.

2 REFERENCIAL TEÓRICO

Neste referencial teórico são apresentados definições e conceitos relacionados aos estudos secundários, mais especificamente, Mapeamento Sistemático e Revisão Sistemática. Também são abordados os tipos de pesquisa e as técnicas de coleta e análise de dados qualitativos, focando, sempre que possível, no contexto de Engenharia de Software (ES).

2.1 ESTUDOS SECUNDÁRIOS

O crescimento do interesse dos pesquisadores de ES na realização de estudos baseados em evidências se deu a partir de práticas do campo da medicina baseada em evidências que se revelou um importante mecanismo para o avanço científico na área (Kitchenham *et al.*, 2006; Jorgensen *et al.*, 2007).

Apesar de já ter sido utilizada em sistemas de informação (Webster e Watson, 2002), a utilização de estudos baseados em evidências para Engenharia de Software (ES) foi primeiramente abordado por Kitchenham que destacou a importância desta abordagem para o desenvolvimento de pesquisas na área (Kitchenham *et al.*, 2004). A introdução do conceito de Engenharia de Software Baseada em Evidências

(ESBE) se deu com Dyba *et al.* (2005), e sua relevância foi logo reconhecida por outros autores da área (Bailey *et al.*, 2007; Budgen *et al.*, 2008; Petersen *et al.*, 2008; Condori-Fernandez *et al.*, 2009; Afzal *et al.*, 2008; Novais *et al.*, 2013).

Segundo Mafra *et al.* (2006), a utilização intensa e sistemática das abordagens baseadas em evidências, poderiam responder a diversas perguntas de Engenharia de Software sobre recomendação de tecnologias mais adequadas para realizar tarefas, custos de implementação de projetos de software, direcionamento de pesquisas, dentre outras, que apoiariam as tomadas de decisão dos profissionais e pesquisadores. Kitchenham *et al.* (2004) diz que com a maior utilização da ESBE, diversos *stakeholders* seriam beneficiados, sejam eles profissionais ou organizações, que usufruiriam do conhecimento científico apresentado nesses estudos. Os usuários também se beneficiariam de soluções e ideias mais precisas para resolução de problemas ou otimização de tarefas.

A ESBE parte das evidências adquiridas através do desenvolvimento de estudos experimentais (Travassos *et al.*, 2007), que ganharam espaço na área de Engenharia de Software a partir dos trabalhos seminais desenvolvidos por Victor Basili nas décadas de 80 e 90 sobre importância da experimentação nesta área de conhecimento (Basili, 2006).

Os estudos experimentais produzem evidências científicas que embasam opiniões ou especulações, muitas vezes construídas de forma não científica pelos desenvolvedores e profissionais da área. Essa abordagem constrói conhecimento a partir de estudos bem definidos, com metodologias confiáveis que podem ser replicados por outros pesquisadores (Perry *et al.*, 2000; Kitchenham *et al.*, 2004). O rigor metodológico agrega valor científico aos estudos e estabelecem estruturas e padrões de pesquisa que podem ser reusadas e evoluídas em pesquisas futuras, e permite a construção do que chamamos de famílias de experimentos (Basili *et al.*, 1999).

Estudos experimentais podem ser classificados como primários e secundários (Yamashita e Moonen, 2013). Os estudos primários conduzem as pesquisas que visam caracterizar uma determinada tecnologia dentro de um contexto específico a partir da coleta e análise direta de novos dados sobre a mesma (Biolchini *et al.*, 2007; Wohlin *et al.*, 2000). Eles contribuem para o aumento do conhecimento científico através do aperfeiçoamento de ferramentas, métodos e técnicas, e da produção de novas

hipóteses (Biolchini *et al.*, 2007). Experimentos (Easterbrook *et al.*, 2008), estudos de caso (Easterbrook *et al.*, 2008) e *surveys* são exemplos destes estudos.

Já os estudos secundários são meios de identificar, avaliar e analisar estudos já existentes, disponíveis e relevantes, sobre uma área, tópico ou objeto de interesse (Mafra *et al.*, 2006). Esses tipos de estudo utilizam resultados já existentes para sintetizar novos resultados através de um processo sistemático e estruturado da pesquisa.

Estudos secundários possibilitam a construção de novas evidências, ou identificam lacunas de evidências, a partir da análise abrangente da evidência já existente sobre um tema de interesse (Mafra *et al.*, 2006; Budgen *et al.*, 2008). Através deles pesquisadores podem ganhar novos conhecimentos ou guiar estudos para lacunas existentes de pesquisa. Este tipo de abordagem também é muito útil para entender o avanço de uma área ou tema de interesse (Petersen *et al.*, 2008; Perry *et al.*, 2008; Biolchini *et al.*, 2007; Kitchenham *et al.*, 2007; Oates *et al.*, 2009) e entende-se que quanto mais estudos deste tipo, maior o conhecimento sintetizado em cada área. Além disso, estudos secundários influenciam melhores tomadas de decisão quanto ao futuro de investigações e práticas em uma determinada área de interesse.

Os estudos secundários podem ser divididos em dois tipos, a saber: Revisões Sistemáticas e Mapeamentos Sistemáticos. Eles serão discutidos a seguir.

2.2 REVISÃO SISTEMÁTICA DA LITERATURA

Dentre as razões mais comuns para realizar um estudo secundário estão: (i) sintetizar novas evidências, a partir das existentes dos beneficiários e limitações de um assunto específico; (ii) resumir evidências e identificar lacunas de pesquisa, melhorando as tomadas de decisão quanto as novas áreas de investigação e; (iii) direcionar novas atividades de pesquisa. As Revisões Sistemáticas (RS) são fundamentais para atingir o primeiro destes objetivos, sintetizar novas evidências.

As RS são estudos secundários realizados através de resultados de estudos primários relevantes de uma determinada área de interesse e são guiados de forma

imparcial e repetível por uma questão de pesquisa específica (Biolchini *et al.*, 2007a; Dyba *et al.*, 2007; Kitchenham *et al.*, 2007). Este tipo de estudo tem aumentado ao longo do tempo em ES (Kitchenham *et al.*, 2009; Kitchenham *et al.*, 2010; Da Silva *et al.*, 2011) devido seu objetivo de identificar, extrair, verificar e analisar todas as evidências disponíveis relacionadas a uma determinada questão de pesquisa e por fornecer resultados abrangentes e detalhados sobre o assunto que está sendo pesquisado (Kitchenham *et al.*, 2007; Kitchenham *et al.*, 2007a; Biolchini *et al.*, 2005).

Em Engenharia de Software a RS é uma ferramenta metodológica significativa no apoio à exploração da validade científica das afirmações feitas em campo, e permite entender melhor a abrangência e as limitações dos métodos empregados no desenvolvimento de software.

A condução de uma RS não é tarefa fácil e a introdução de diretrizes específicas para cada área contribuem para a qualidade dos estudos realizados (Biolchini *et al.*, 2007; Kitchenham *et al.*, 2007), desta forma, Kitchenham *et al.* (2007) desenvolveram as orientações para condução de uma RS.

Essas diretrizes foram adequadas às necessidades da comunidade de ES da época, visto o estágio inicial da disciplina dez anos atrás, mas elas têm evoluído com o passar do tempo. Principalmente no que se refere à síntese de dados para produção de novas evidências.

As diretrizes básicas para uma RS seguem as etapas descritas na Figura 1. Elas são divididas em três fases: (1) Planejamento da Revisão; (2) Execução da Revisão; e (3) Análise da Revisão (Biolchini *et al.*, 2007; Kitchenham *et al.*, 2007; Pai *et al.*, 2003). A boa documentação dessas etapas permite a replicação ou evolução do estudo realizado. Uma boa revisão deve não somente permitir a sua replicação, mas também deve facilitar que outros pesquisadores realizem futuras atualizações na sua execução a partir da disponibilização de novas evidências e estudos científicos na literatura.

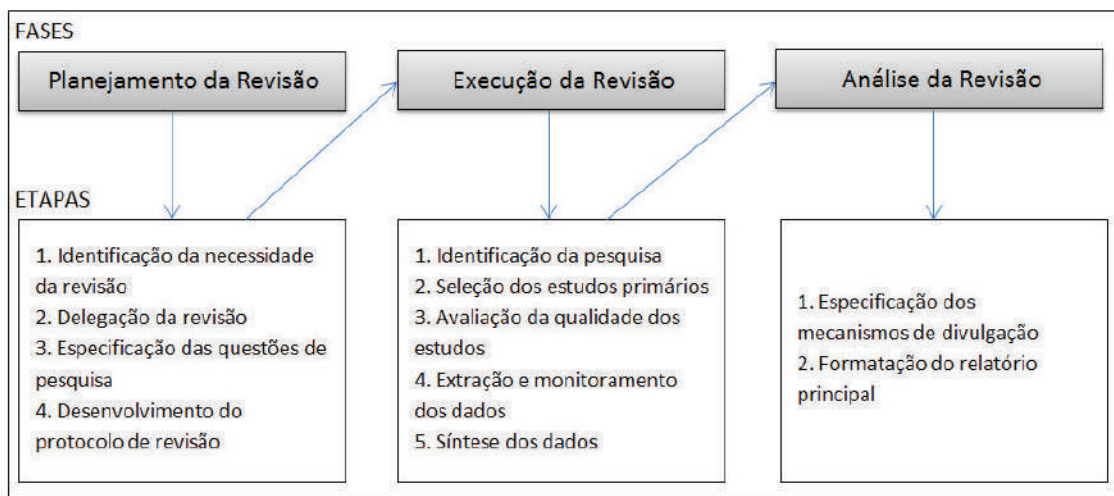


Figura 1: Etapas de uma revisão sistemática (Kitchenham *et al.*, 2007)

Na Figura 1 as etapas de avaliação do protocolo da revisão, presentes na fase de planejamento da revisão e a de avaliação do relatório, presente na fase de análise da revisão, não foram contempladas por não serem obrigatórias em uma RS segundo (Kitchenham *et al.*, 2007).

Abaixo descrição resumida de todas as fases de uma revisão sistemática segundo Kitchenham *et al.* (2007) com suas respectivas etapas.

2.2.1 Planejamento da Revisão

Na fase de planejamento da revisão, um plano é traçado para listar os objetivos, direcionamentos e estratégias a serem adotadas na RS. Esta fase contempla as seguintes etapas:

- (1) **Identificação da necessidade da revisão** - Nesta etapa verifica-se, por exemplo, se a RS é decorrente de uma exigência dos pesquisadores para fazer alguma verificação de um assunto ou prelúdio para outras atividades de pesquisa;
- (2) **Delegação da revisão** - Utilizada quando a organização delega a pesquisadores a realização de uma RS de um tema que ela não possui

tempo nem experiência para realizar. Ela é feita através de um documento de delegação que especifica o trabalho necessário e encaminhado a grupos de pesquisa que queiram realizar a revisão;

- (3) **Especificação das questões de pesquisa** - É uma das etapas mais importantes de uma RS, pois, é ela que orienta toda a sua metodologia incluindo o processo de pesquisa, extração e análise dos dados que devem estar alinhados ao objetivo de responder as questões de pesquisa definidas nesta etapa (Counsell, 1997);
- (4) **Desenvolvimento do protocolo da revisão** - Etapa onde os métodos que serão utilizados para conduzir a RS são especificados, ou seja, onde são definidos os planos básicos que orientam e definem claramente como deve ser realizada a revisão. Critérios de inclusão e exclusão de estudos, estratégias de pesquisa, cronograma do projeto, estratégia de extração dos dados, dentro outros, são devidamente documentados neste protocolo.

2.2.2 Execução da Revisão

Na fase de execução da revisão, todos os passos descritos no protocolo são executados e devem ser seguidos rigorosamente. Ela contempla as seguintes etapas:

- (1) **Identificação da pesquisa** - Etapa onde a estratégia de busca para identificar os estudos potencialmente relevantes para a pesquisa é aplicada. Esse processo de busca é rigoroso e inerente das RS e composto das atividades de: (i) geração de uma estratégia de busca; (ii) identificação dos vieses das publicações; (iii) gerenciamento da bibliografia e obtenção de artigos; e (iv) documentação da pesquisa;
- (2) **Seleção dos estudos primários** - Nesta etapa os estudos primários são avaliados quanto a sua real importância para a pesquisa, através dos critérios de inclusão e exclusão estabelecidos no protocolo da revisão que podem ser refinados durante o processo (Kitchenham *et al.*, 2004).

A qualidade da seleção dos estudos é fundamental. Ela é geralmente realizada por mais de um pesquisador e os conflitos são resolvidos utilizando reuniões de consenso (Da Silva *et al.*, 2011; Brereton *et al.*, 2007). Também, pode-se aplicar a análise de concordância – *kappa* (Cohen *et al.*, 2006), ou seja, as incertezas são investigadas por uma análise de sensibilidade (Kitchenham *et al.*, 2004). Outro método é o teste de confiabilidade entre avaliadores que utilizam uma amostra aleatória dos estudos em que um revisor secundário aplica os critérios de inclusão e seleção definidos no protocolo e deve ter como resultado os mesmos estudos obtidos pelo revisor primário, sem diferenças, garantindo a concordância (Khan *et al.*, 2011);

- (3) **Avaliação da qualidade dos estudos** - Nesta etapa, além dos critérios de inclusão e exclusão, a qualidade dos estudos primários deve ser avaliada para estimar a sua importância na síntese e direcionamento da compreensão dos dados e recomendações para pesquisas futuras (Hamer e Gill, 2005). Dyba *et al.* (2008) propôs uma lista de critérios de avaliação da qualidade dos estudos primários específica para a área de Engenharia de Software que estabelecem através de valores booleanos (1 verdadeiro e 0 falso) uma classificação para medir a confiança dos resultados de estudos de RS na área;
- (4) **Extração e monitoramento dos dados** - Etapa onde os dados dos estudos primários selecionados são extraídos e registrados com o objetivo de obter dados necessários para abordar as questões de pesquisa e os critérios de qualidade do estudo. Informações dos estudos como: título, autor, local de publicação, detalhes da publicação, dentre outros, devem ser obtidas na extração. Os dados devem ser registrados em formulários de extração de dados definidos no desenvolvimento do protocolo da revisão;
- (5) **Síntese dos dados** - Etapa onde os dados coletados dos estudos primários selecionados são agrupados, resumidos e expandidos através da análise combinada dos estudos primários. Essa síntese pode ser descritiva e/ou quantitativa e é uma das fases mais críticas da Revisão

Sistemática, pois busca realçar semelhanças e diferenças dos resultados dos estudos primários com o objetivo de compreender a associação entre os mesmos e apresentar respostas às questões de pesquisa da RS com a produção de evidências.

2.2.3 Análise da Revisão

A análise da revisão é a última fase da RS, onde são apresentados os resultados e realizada a divulgação da avaliação. Nela estão presentes as etapas de:

- (1) **Especificação dos mecanismos de divulgação** - Etapa onde é realizada a publicação dos resultados da RS de forma eficaz, em páginas da internet, conferências, periódicos ou revistas, dentre outros. Recomenda-se que as estratégias de divulgação sejam devidamente definidas na etapa de desenvolvimento do protocolo da revisão ou, caso haja, na etapa de delegação da revisão;
- (2) **Formatação do relatório principal** - Nesta etapa a formatação do relatório, geralmente, formatado como relatório técnico, seção de uma tese de doutorado, periódico ou artigo para uma conferência é realizada.

2.3 MAPEAMENTO SISTEMÁTICO

O Mapeamento Sistemático (MS) é uma abordagem que classifica e estrutura evidências de pesquisas, referentes a uma área de interesse, visando responder ou contestar questões de investigação (Budgen *et al.*, 2008; Petersen *et al.*, 2008). Ele se diferencia das Revisões Sistemáticas, pois sumariza os dados coletados, mas não sintetiza seus resultados em busca de novas evidências dos mesmos.

Esse método de estudo secundário é bastante utilizado para obter uma visão geral e resumida de estudos primários de áreas que se deseja verificar a deficiência de estudos (Kitchenham *et al.*, 2007). Nele a classificação e caracterização dos resultados serão apresentadas em um nível de granularidade maior, com o objetivo de responder as questões de pesquisa e possibilita a identificação de lacunas de pesquisa.

O MS é o método mais indicado para realizar atividade em um domínio de investigação em que o tema é muito amplo ou existe uma probabilidade baixa de encontrar evidência, desta forma recomenda-se a sua execução antes da realização de uma RS (Kitchenham *et al.*, 2007).

Segundo Arksey *et al.* (2005) e Budgen *et al.* (2008) o mapeamento sistemático justifica-se por ser uma revisão útil para mapear e apresentar lacunas de pesquisa de uma grande quantidade de estudos disponíveis, de maneira imparcial, o que contribui para a comunidade de pesquisa com uma síntese confiável de dados.

Com o MS são apresentados resultados resumidos e de fácil visualização, com base em evidências encontradas nos estudos, o que apoia o planejamento de novas pesquisas e evita as duplicações de estudos e retrabalho dos pesquisadores (Arksey *et al.*, 2005; Budgen *et al.*, 2008).

2.3.1 Processo do Mapeamento Sistemático

Enquanto o processo de uma RS é mais aprofundado na análise dos resultados dos estudos primários com três fases com várias etapas, no MS essas fases visam somente classificar os estudos.

Budgen *et al.* (2008) descreve um MS em três etapas também chamadas de estágios iniciais de um estudo secundário. São elas: (i) a etapa da procura onde devem ser identificados os estudos primários de interesse da pesquisa; (ii) a etapa de inclusão/exclusão onde são selecionados somente os estudos que contribuem para responder as questões de pesquisa e; (iii) a etapa de viés/validade onde são realizadas as avaliações de qualidade dos resultados produzidos.

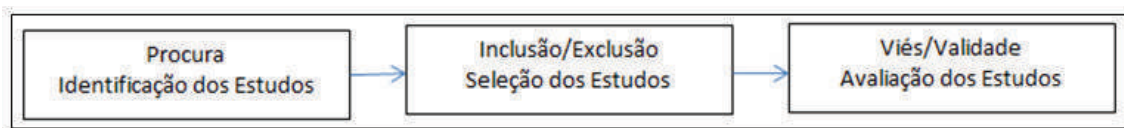


Figura 2: Etapas iniciais do mapeamento sistemático

Essas etapas podem eventualmente ser seguidas pela etapa de extração e análise dos dados (síntese de resultados) de uma revisão sistemática.

Petersen *et al.* (2008) apresenta um processo mais detalhado de como realizar um estudo de mapeamento sistemático, que foi adotado pela comunidade de Engenharia de Software como processo padrão para a realização destes tipos de estudos na área.

O processo apresenta as seguintes etapas: (1) definição de questões de pesquisa; (2) realização da busca de estudos relevantes; (3) triagem dos estudos; (4) resumo de palavras-chave utilizadas e; (5) extração de dados e o processo de mapeamento. A Figura 3 apresenta o processo de mapeamento sistemático descrito por Petersen *et al.* (2008).

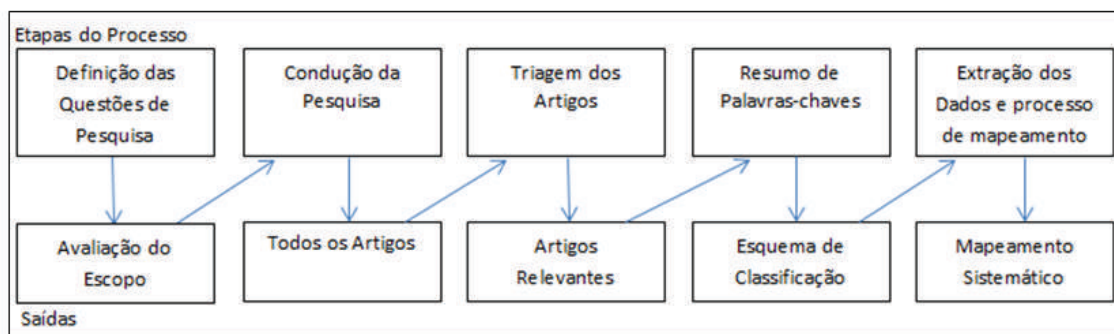


Figura 3: Processo do mapeamento sistemático (Petersen *et al.*, 2008)

Petersen *et al.* (2008) e Condrón-Fernandes *et al.* (2009) não abordam a etapa de definição do protocolo em seus processos para realização de um mapeamento sistemático, visto que este é considerado um artefato não obrigatório nesse tipo de estudo. Em controversa, outros pesquisadores defendem que um protocolo detalhado é de fundamental importância para diminuir o viés e aumentar a probabilidade de replicação do MS (Dyba *et al.*, 2007; Berry *et al.*, 1997).

Abaixo é apresentada a descrição resumida de todas as etapas do processo de MS apresentado por Petersen *et al.* (2008).

2.3.2 Definição das questões de pesquisa

As questões de pesquisa são importantes desde a sua construção, até a sua resposta, visto que quando definidas de forma clara melhoram a probabilidade de resultados de alta qualidade e a sua precisão leva a conclusões importantes e definem o escopo do mapeamento.

Nesta etapa as questões de pesquisa utilizadas para direcionar o MS devem ser definidas junto à razão para a sua inclusão. Muitas vezes elas têm o objetivo de mapear as frequências de publicações durante os anos, para verificar as tendências de uma área de conhecimento. Sugere-se que essas questões sejam elaboradas segundo os critérios PICOC (População, Intervenção, Comparação, Resultados e Contexto) (Petticrew *et al.*, 2008; Kitchenham *et al.*, 2007).

2.3.3 Condução da pesquisa

As bases de dados científicas são bibliotecas digitais ou sites da internet onde os estudos primários são armazenados de forma estruturada ou semiestruturada. Dentre elas temos: *IEEE Xplore*, *ACM Digital Library*, *Springer*, sites de conferências, dentre outras. Segundo Petersen *et al.* (2008), os estudos primários são identificados através de *strings* de busca executadas nas bibliotecas digitais ou através de buscas manuais em conferências ou periódicos.

Estas *strings* são, geralmente, palavras-chave combinadas que se utilizam de conectores como OR ou AND, relacionadas com as questões de pesquisa adotadas. Elas

são refinadas a depender dos resultados inicialmente obtidos. Kitchenham *et al.* (2007) aborda boas práticas para se criar uma boa *string* de busca ou sequência de pesquisa.

Nesta etapa de condução da pesquisa são definidas quais bases de dados científicas e estratégias de busca serão utilizadas. E como resultado todos os estudos primários a serem processados na pesquisa são identificados e precisam ser avaliados quanto a sua real relevância para o estudo.

2.3.4 Triagem dos artigos

Nesta etapa é realizada a seleção dos estudos encontrados que atendem aos critérios de inclusão e exclusão pré-definidos para o mapeamento. Esses critérios de seleção também são influenciados pelas questões de investigação, utilizados para incluir ou excluir artigos que não são relevantes para responder as questões de pesquisa e aplicados com o objetivo de melhorar os resultados obtidos (Kitchenham *et al.*, 2007). Desta forma, eles devem ser definidos durante a definição das questões de pesquisa e refinados ao longo do processo de busca dos artigos.

Ao final desta etapa os artigos relevantes para a pesquisa estarão selecionados para análise.

2.3.5 Resumo das palavras-chave

Nesta etapa é definido o esquema de classificação dos artigos. Petersen *et al.* (2008) propõe classificar os estudos em facetas conforme o processo sistemático descrito na Figura 4.

Neste processo o pesquisador lê os resumos dos artigos em busca de palavras-chave que reflitam a sua contribuição para a pesquisa, bem como, o seu contexto. Caso os resumos não sejam de uma boa qualidade, a introdução ou conclusão

dos artigos podem ser também analisadas. As palavras-chaves encontradas são refinadas com o objetivo de melhorar a classificação e aumentar o entendimento sobre a natureza e contribuição da pesquisa. Desta forma, os revisores buscam desenvolver um conjunto de categorias representativo da população estudada. Com o conjunto final de palavras-chave, elas são utilizadas para formar as categorias para o mapa sistemático.

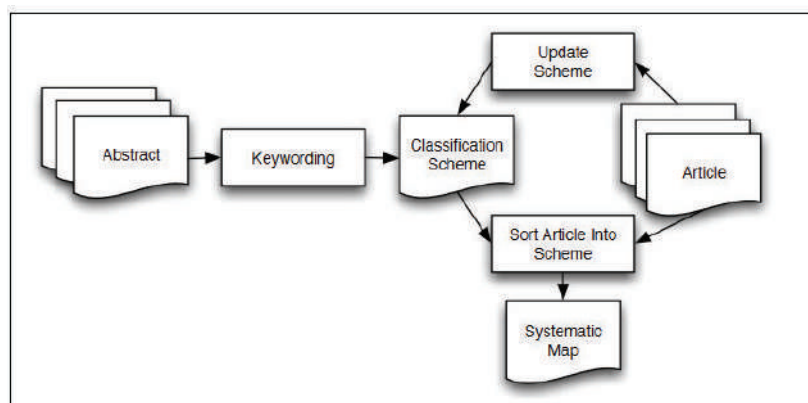


Figura 4. Plano de classificação (Petersen *et al.*, 2008)

Como forma de minimizar o esforço desta etapa, esquemas de classificação produzidos por outros autores também podem ser utilizados desde que seja observada a área de conhecimento e especificações de cada pesquisa (Jorgensen *et al.*, 2007; Kagdi *et al.*, 2007).

2.3.6 Extração dos dados e o processo de mapeamento

Nesta fase são realizadas as leituras completa de todos os artigos selecionados e sua classificação no esquema desenvolvido anteriormente, ou seja, a extração dos dados relevantes para a pesquisa é iniciada. Os dados extraídos podem ser registrados em tabelas eletrônicas, o que documenta este processo e permite posteriores medições das frequências de publicações em cada categoria.

Segundo Petersen *et al.* (2008) nesta fase também ocorre a evolução do esquema de classificação, visto que, novas categorias surgem ou categorias existentes que são semelhantes se unem, o que melhora o esquema.

As análises dos resultados das avaliações das frequências/quantidades das publicações por categoria dentro do esquema também ocorrem nesta fase. Esses resultados podem ser apresentados com a utilização, por exemplo, de gráficos de bolhas. Segundo Petersen *et al.* (2008) o gráfico de bolhas suporta melhores análises do que tabelas de frequência.

Durante este processo são identificadas lacunas e novas oportunidades de pesquisa.

2.4 DIFERENÇAS ENTRE MAPEAMENTO SISTEMÁTICO E REVISÃO SISTEMÁTICA

Em seus estudos Budgen *et al.* (2008) abordam as diferenças e Petersen *et al.* (2008) faz uma comparação de RS e MS. Resumidamente RS e MS possuem as seguintes características:

- **Revisões Sistemáticas:**
 - Questões de pesquisa mais definidas e focadas;
 - Extração de dados mais detalhada e direcionada a identificação de melhores práticas e efetividade da área de estudo;
 - Apresenta avaliação de qualidade dos estudos primários, com o objetivo de reduzir a probabilidade de falsos resultados.
- **Mapeamentos Sistemáticos:**
 - Questões de pesquisa amplas, exploratórias e descritivas sobre tendências de pesquisa;
 - Síntese de dados e classificação dos artigos com o objetivo de responder as questões de pesquisa sem ser uma tarefa demorada;
 - Extração de dados abrangente e focada na classificação e categorização dos resultados;

- Termos de busca menos focados no assunto de pesquisa, por isso há a probabilidade do retorno de um número muito grande de estudos;
- Resultados representados graficamente o que traz uma melhor visualização das informações.

A principal diferença entre os mapeamentos sistemáticos e as revisões sistemáticas é que estas últimas buscam a síntese dos resultados primários através das análises das evidências encontradas nos estudos, o mapeamento sistemático não compartilha desta característica, pois seu processo não avalia os estudos com tal profundidade (Petersen *et al.*, 2008). Por esta razão a principal distinção entre RS e MS se dá no domínio e procedimentos que são utilizados nas suas análises, visto que em um MS é mapeado principalmente a pesquisa por diferentes classificações dos estudos primários incluído no estudo secundário, enquanto a RS fornece uma síntese do conhecimento em uma área relacionada a uma questão de pesquisa específica.

Apesar de ambos possuírem a característica de identificar lacunas de pesquisa, estes estudos também se diferenciam em seu escopo, pois em um MS ele é mais amplo do que em uma RS.

Os procedimentos de seleção dos estudos primários que são incluídos durante a execução dos estudos secundários são os mesmos para RS e MS. A diferença se dá na estratégia de busca, visto que a RS deve ser mais rigorosa nos critérios de inclusão de estudos primários para formar uma base sólida para geração de evidências (Kitchenham *et al.*, 2011). Este critério de rigor está relacionado, por exemplo, as limitações referentes às fontes de dados utilizadas para extrair os artigos relevantes.

A etapa de síntese deve ser parte integrante da RS (Cruzes *et al.*, 2011). O MS é um estudo sem essa etapa e, muitas vezes, aborda uma área em vez de uma questão de pesquisa mais específica. Por isso, um MS pode ser caracterizado como o primeiro passo a ser tomado antes de se realizar uma RS, um ponto de partida para futuras pesquisas (Kitchenham *et al.*, 2011) ou um complemento da RS (Petersen *et al.*, 2008).

Percebe-se que os mapeamentos sistemáticos visam mapear a área de interesse de forma abrangente, enquanto as revisões sistemáticas focam em responder com detalhes as questões de pesquisa escolhidas.

Quanto aos resultados, um MS é mais resumido e seus resultados são mais simples de visualizar do que em uma RS, o que apresenta uma visão geral de uma área de conhecimento e identifica seus *gaps*.

Diante do exposto, e observadas às diferenças entre RS e MS, os pesquisadores devem avaliar o contexto e decidir qual o melhor processo a ser adotado para as suas pesquisas.

Depois de escolher qual processo a ser utilizado, os pesquisadores poderão analisar os dados com a utilização de abordagens metodológicas de forma qualitativa e/ou quantitativa. Na próxima seção, esses tipos de pesquisa são caracterizados e discutidos resumidamente.

2.5 TIPOS DE PESQUISA

A pesquisa se apresenta como um esforço sistemático para descobrir novas informações e/ou relações e para verificar e ampliar o conhecimento existente. Essas descobertas se dão pela aplicação de pesquisas de investigação científica, que quanto a sua abordagem, são subdivididas em dois tipos, quantitativas e qualitativas (Godoy, 1995).

No primeiro tipo sua principal função é apresentar dados, indicadores e tendências numericamente analisáveis, visto que, seus resultados devem evidenciar a frequência dos eventos (Chizzotti, 2006, Richardson, 2008) e preocupar-se com a sua medição objetiva e quantificável (Godoy, 1995). Desta forma, essa abordagem considera que tudo pode ser transformado em dados numéricos, opiniões e informações para que uma melhor análise e avaliação possam ser realizadas.

Na pesquisa quantitativa, uma amostra representativa da população estudada assegura a possibilidade de uma generalização dos resultados (Gunther, 2006). Frequentemente elas utilizam a formulação de hipóteses e classificação da relação existente entre as variáveis (Cook e Reichardt, 1979).

Por outro lado, o tipo de pesquisa qualitativa caracteriza-se pela ausência de medidas numéricas e destaca-se pela escrita tanto para obter dados como para

divulgação dos resultados do fenômeno estudado que pode ser ações dos indivíduos, grupos ou organizações em seu ambiente ou contexto social, com o foco na observação, descrição, compreensão e significado do fenômeno pesquisado (Godoy, 1995). Ela procura analisar características mais profundas e subjetivas do assunto em estudo e considera que há um vínculo ativo entre o mundo real e o sujeito que não pode ser facilmente traduzido em números. Três tipos são bastante conhecidos: a pesquisa documental, o estudo de caso e a etnografia (Godoy, 1995).

De forma geral, os métodos qualitativos são menos estruturados, proporcionam um relacionamento mais longo e flexível entre o pesquisador e os entrevistados, e lidam com informações mais subjetivas, amplas e com maior riqueza de detalhes do que os métodos quantitativos. Nesta abordagem os dados são apresentados pelos pesquisadores de forma descritiva e as análises tendem a serem feitas de forma indutiva, com o foco no processo e nos seus resultados (Cook e Reichardt, 1979).

Tanto as abordagens quantitativas quanto as qualitativas serão utilizadas nesta dissertação, visto que é preciso analisar dados diversos e transformar resultados em informações de fácil leitura para os pesquisadores, sejam elas numéricas ou textuais.

A próxima seção aborda as técnicas que podem ser utilizadas para coletar dados qualitativos. Um enfoque maior foi dado à técnica de entrevista, processo que compõe o método de pesquisa utilizado nesta dissertação.

2.5.1 Técnicas de Coleta de Dados Qualitativos

A coleta de dados é utilizada para pesquisar, reunir documentos e provas e buscar dados de um ou vários temas na procura de informações que facilitem uma posterior análise e conclusão sobre o assunto. Para que um trabalho seja executado ou um fato ou fenômeno seja esclarecido a coleta de dados ajuda na análise e é sistematicamente registrada e planejada (Gil, 2002).

Diversas técnicas podem ser utilizadas para a realização de uma coleta de dados. Essas técnicas são um conjunto de processos utilizados por uma ciência na

obtenção de seus propósitos (Lakatos e Marconi, 2010). Cada uma delas tem seus próprios padrões, melhores práticas e regras.

De acordo com Lakatos e Marconi (2010) entre as técnicas para realização de coleta de dados temos:

- **Entrevista:** Comunicação verbal, entre duas pessoas, de forma metódica, com o objetivo de coletar informações sobre um determinado assunto;
- **Observação:** É realizada por pesquisadores em contato direto com os fatos ou fenômenos que se deseja estudar para obter informações sobre aspectos da realidade;
- **Questionário:** Instrumento que coleta dados por meio de perguntas respondidas por escrito pelo pesquisado.
- **Formulário:** É composto de perguntas apresentadas por escrito pelo pesquisador e respondidas de forma oral pelo pesquisado;
- **Medidas de opinião e de atributos:** Forma de observação padronizada que visa assegurar a proporção entre diferentes opiniões e atitudes, com o objetivo de compará-las;
- **História de vida:** Obtenção de dados das experiências vividas por uma pessoa que tenha um significado importante para o conhecimento do objeto de estudo.

2.5.2 Entrevistas

Com as entrevistas se pressupõe que se pode entender como o mundo é através das respostas a perguntas abertas e estruturadas sobre as experiências de um entrevistado. Por coletar informações através da comunicação verbal entre pesquisador e pesquisado sobre um tema científico, é a técnica mais usada em atividades de trabalho de campo (Minayo, 2008).

Esta técnica tem o objetivo de mapear e compreender determinado contexto onde os entrevistados estão envolvidos, ou seja, ela produz elementos básicos para que

de forma detalhada haja uma compreensão das práticas, crenças, atitudes, valores e motivação em relação aos atores sociais e contextos sociais específicos (Minayo, 2008; Duarte, 2004).

Segundo Lakatos e Marconi (2010), ao utilizar-se da entrevista tem-se a oportunidade de conversar face a face com o entrevistado, o que possibilita: (i) verificar acontecimentos ocorridos; (ii) saber a opinião das pessoas sobre os acontecimentos; (iii) saber o sentimento da pessoa sobre o fato ou seu significado para ela; (iv) descobrir quais foram, são ou seriam as condutas das pessoas, sejam elas passadas, presentes ou planejadas; e (v) descobrir fatores que influenciam os pensamentos, sentimentos ou ações das pessoas.

Segundo Duarte *et al.* (2004), para que uma boa entrevista seja realizada, são exigências para o pesquisador: (i) ter objetivos de pesquisa bem definidos; (ii) ter conhecimento do contexto de investigação, com alguma profundidade; (iii) fazer um piloto de entrevista com o roteiro definido, para evitar “engasgos” durante as entrevistas válidas; e (iv) ter algum nível de informalidade, sem perder os objetivos da pesquisa.

2.5.2.1 Tipos de Entrevistas

As entrevistas se diferem quanto ao propósito do investigador, mas, em todos os casos, os entrevistados descrevem suas próprias experiências de forma detalhada e incluem narrativas pessoais ou histórias de vida (Shah *et al.*, 2006). Quanto a essa estrutura, ou forma de organização, elas são classificadas em três tipos ou categorias. Neste contexto, adotou-se a classificação usada por Nunes (2007):

- **Estruturadas:** Utiliza-se de um roteiro previamente estabelecido com perguntas predeterminadas. As perguntas não podem ser adaptadas à situação.
- **Semiestruturadas:** Utiliza-se de perguntas parcialmente formuladas que vão sendo adequadas à situação durante a entrevista.

- **Não estruturadas:** As perguntas são livres sem prévia formulação e seguem em qualquer direção considerada pelo pesquisador adequada. Busca explorar amplamente a questão durante a entrevista.

Os tipos de entrevistas mais utilizadas em pesquisas qualitativas são as não estruturadas e as semiestruturadas, que se diferenciam pelo grau de participação representado pelo pesquisador.

Independente da modalidade de entrevista escolhida deve-se fazer um planejamento prévio e conhecer algumas especificidades do uso desta técnica. Segundo Minayo (2008) e Lakatos e Marconi (2010) o pesquisador deve tomar os seguintes cuidados antes e durante esse processo: (1) Planejar com antecedência a entrevista; (2) Fazer a seleção dos entrevistados garantindo a sua representatividade para a pesquisa; (3) Conseguir conhecimento prévio sobre o entrevistado; (4) Marcar com antecedência o local, o horário de início e a duração prevista das entrevistas; e (5) Criar condições confortáveis e apropriadas para a realização da entrevista.

Os dados coletados nas entrevistas são submetidos a análises para que informações relevantes possam ser descobertas. Na próxima seção são apresentadas algumas técnicas para análise de dados qualitativos.

2.6 TÉCNICAS PARA ANÁLISE DE DADOS QUALITATIVOS

A análise de dados é considerada como o núcleo central da pesquisa qualitativa, pois, evidencia a ideia de que os dados qualitativos têm significado e devem ser explicados em análise, não apenas para apresentar a diversidade de temas abordados pelas pessoas, mas também para reconhecer e analisar as formas como suas comunicações são enquadradas e modeladas por elas (Gibbs, 2009).

Os dados adquiridos em abordagens quantitativas que contam com o apoio de softwares estatísticos, estatística descritiva e multivariada e teste de hipóteses são analisados de forma diferente dos dados que são coletados nas pesquisas de abordagens qualitativas. Para esse contexto, são utilizadas as técnicas de análise de dados qualitativos.

Nesses tipos de abordagem objetiva-se extrair significado dos dados de textos e imagens, desta forma, aspectos como manipulação e interpretação de dados são utilizados (Creswell, 2010; Gibbs, 2009; Creswell, 2007), ou seja, essas técnicas são interpretativas e procuram descrever, decodificar e traduzir para chegar a um significado do fenômeno estudado.

A depender do tipo de dado as análises qualitativas são mais gerais ou mais específicas, mas, se assemelham no fato de serem baseadas em análises textual, desta forma, precisam ser preparados e analisados como texto. Duas atividades são envolvidas na análise qualitativa: (i) o desenvolvimento de uma consciência dos tipos de dados que podem ser examinados, escritos e explicados e (ii) o desenvolvimento de uma sequência de atividades práticas adequadas aos tipos e quantidade de dados que devem ser examinados (Gibbs, 2009).

Segundo Minayo (2008), as abordagens mais conhecidas para análise de dados quantitativos são: Análise de conteúdo, análise dialética/hermenêutica e análise do discurso. Na Engenharia de Software métodos como: metaetnografia (Campbell *et al.*, 2003), metassíntese (Paterson, 2001), síntese temática (Cruzes e Dyba, 2011a; Cruzes e Dyba, 2011b), dentre outros, são utilizados para análise e síntese de dados qualitativos. Não é o foco deste trabalho descrever todos esses métodos, somente a síntese temática, que se utiliza de análise temática. Este método é descrito na próxima seção, pois foi o método utilizado para análise dos dados qualitativos desta dissertação.

2.6.1 Síntese Temática

Síntese temática é a expressão utilizada por Thomas e Harden (2008) para apresentar este método de análise e síntese de dados qualitativos baseado nos princípios da análise temática. A análise temática é utilizada para identificar, analisar e retornar padrões (temas) de dados qualitativos (Braun e Clarke, 2006). É um dos métodos de análise qualitativa mais utilizados na ES (Cruzes e Dyba 2011a). Ela descreve os dados em detalhes e interpreta aspectos do assunto pesquisado e também pode ser um método

essencialista ou realista para relatar experiências, significados e a realidade dos participantes (Braun e Clarke, 2006).

Neste contexto, a síntese temática foi formalizada por Thomas e Harden (2008) por meio de três etapas sistemáticas (codificação, organização dos códigos para a construção de temas e desenvolvimento de temas descritos) que combinam e adaptam a metaetnografia e *grounded theory*.

Cruzes e Dyba (2011a) analisaram o método de Thomas e Harden (2008) e desenvolveram a abordagem de síntese temática com o foco em ES. Esse método foi aplicado em dados coletados de diferentes estudos primários, com o objetivo de sumarizar, integrar, combinar e comparar seus resultados.

A Figura 5 representa uma visão geral do processo de síntese temática em ES. Esta figura apresenta os passos do método, sua relação e os resultados obtidos em cada um deles. Estes passos estão descritos a seguir:

1. Realização das leituras iniciais dos dados e/ou textos (geralmente muitas páginas de texto disponíveis para análise);
2. Identificação de segmentos de texto específicos, que são importantes para o objetivo da síntese, tendo como resultado muitos segmentos de texto extraídos.
3. Rotulação dos segmentos de texto identificados na etapa anterior, tendo como resultado a codificação dos rótulos.
4. Redução dos códigos sobrepostos e a tradução de códigos em temas, tendo como resultado os temas de interesse e análise.
5. Criação de um modelo de alto nível que contém os temas e como resultado os temas tratados que representam os dados e/ou textos analisados são obtidos.

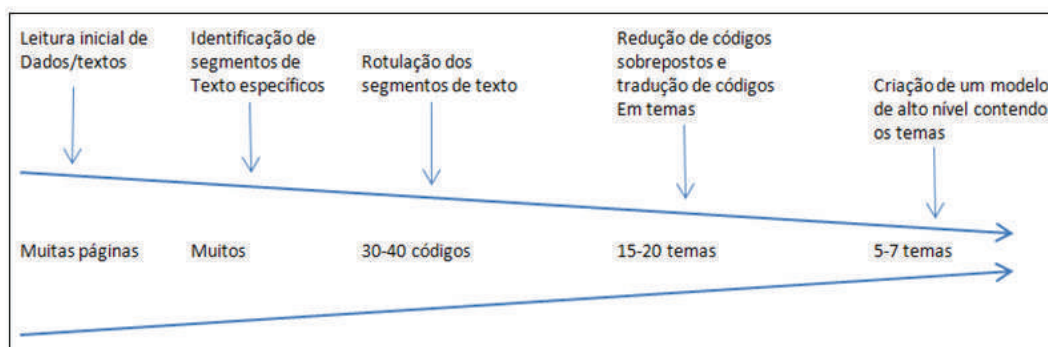


Figura 5. Método de síntese temática (adaptado de Creswell, 2007; Cruzes e Dyba, 2011a)

Cruzes e Dyba (2011a) apresentaram as principais etapas da síntese temática em ES e uma lista de itens (*checklist*), que devem ser checados pelo pesquisador, relacionados a cada etapa do processo. Posteriormente, Prates (2016) utilizou como referência estas etapas e *checklist* e propôs um método refinado de síntese temática com as recomendações por etapa descrito no Quadro 1.

Quadro 1. Refinamento do Método de Síntese Temática (Prates, 2016)

O que fazer? (Etapas)	Como Fazer? (Recomendações)
<p>Extração de Dados (Extrair dados de estudos primários, incluindo informação bibliográfica, objetivos, contexto e resultados.)</p>	<p>1. Definir a adoção ou não de uma ferramenta de extração; 1.1. Caso a decisão seja de realização da extração e codificação de forma manual é necessário garantir a rastreabilidade das decisões tomadas, extrações realizadas, codificações e identificação de temas. 1.2. Para extrações manuais, instrumentos de extração devem ser planejados e gerados para tornar o processo mais consistente, diminuindo a possibilidade de retrabalhos. 2. Realizar encontros iniciais para nivelar o conhecimento, equalizando as decisões tomadas para a realização do trabalho (tema central, critérios de inclusão, exclusão, alinhamento das expectativas em relação às questões de pesquisa); 3. Avaliar a necessidade de um pesquisador especialista no tema central adotado para a síntese temática. Caso os pesquisadores não detenham domínio do tema central para o qual a RS foi realizada, recomenda-se a participação de um especialista, principalmente para promover as verificações dos passos apresentados, aumentando a confiabilidade da síntese; 4. Ler todos os artigos selecionados, buscando imersão no contexto; 5. Analisar e definir o arcabouço de extração; 5.1. O arcabouço de extração sugere as possíveis dimensões que poderão ser codificadas no item seguinte. 5.2. É recomendado que o arcabouço de extração considere, no mínimo, as seguintes dimensões: publicação, contexto e resultados. 6. Analisar e dividir os artigos selecionados para extração, considerando a execução de etapas iterativas de extração e codificação. Recomenda-se a adoção de critérios para esta divisão em etapas; 7. Extrair os dados dos estudos selecionados, observando na primeira etapa as principais seções onde as fontes de informação são comumente encontradas; 7.1. Detalhes das dimensões consideradas no arcabouço de extrações devem ser extraídos dos artigos selecionados; 7.2. Diferenciar durante a extração o que é apenas suposição dos autores do que efetivamente é contribuição dos artigos; 7.3. As principais seções onde as fontes de informações são comumente encontradas tendem auxiliar nas próximas etapas de extração, visto que norteiam na maioria dos casos, onde encontrar as informações que estão sendo consideradas na extração. Além disso, reduzem o esforço de leitura, sem prescindir da sistematização do processo. 7.4. As seguintes perguntas podem auxiliar na identificação de uma constatação: - Indica resultados das medições? - Resumem os dados brutos? - Destacam alguma característica específica dos dados brutos? - Fornecem <i>insights</i> sobre tabelas ou figuras apresentadas? - Resumem os resultados das análises? - Pode ser usado para responder às questões de pesquisa apresentadas? - Refletem os principais resultados do estudo? 7.5. Todos os segmentos de texto que exemplifiquem a mesma ideia devem</p>

	<p>ser extraídos.</p> <p>7.6. Sempre que possível, a extração deve ocorrer de forma independente por dois ou vários pesquisadores, comparando os resultados;</p> <p>7.7. Manter a rastreabilidade entre as informações contidas nos artigos e os segmentos de texto extraídos, caso o processo seja sem o apoio de ferramentas de extração.</p> <p>8. Verificar a extração.</p> <p>8.1. A Verificação deve ser feita por um pesquisador diferente daquele que realizou a extração dos dados.</p> <p>8.2. Na verificação é importante garantir que os itens relevantes para as dimensões consideradas foram extraídos.</p> <p>8.3. É importante combinar com os demais envolvidos na síntese/verificação, os passos que serão adotados na execução das verificações, alinhando os critérios de verificação adotados.</p>
<p>Codificação de Dados (Identificar e codificar conceitos interessantes, categorias, descobertas e resultados de uma forma sistemática em todo o conjunto de dados)</p>	<p>9. Definir as dimensões que serão codificadas;</p> <p>9.1. Como dito no passo anterior, as dimensões que serão codificadas devem estar contidas no arcabouço de extração.</p> <p>10. Definir os métodos de codificação que serão adotados, considerando a natureza da dimensão e sua representatividade nas análises;</p> <p>11. Definir a abordagem para realização da codificação (dedutiva, indutiva, integrada);</p> <p>11.1. A abordagem integrada é uma opção interessante de ser adotada, pois concilia a preocupação de não ser restritiva com o aproveitamento do conhecimento extenso de algum especialista que esteja participando do método.</p> <p>No entanto, deve-se considerar a experiência dos pesquisadores envolvidos e o estágio da pesquisa.</p> <p>12. Manter a rastreabilidade na codificação.</p> <p>13. Codificar os conjuntos de dados, realizando comparações constantes, considerando o conhecimento dos pesquisadores e as questões de pesquisa. Esta codificação deve considerar a rastreabilidade definida para a síntese temática. Caso sejam identificados códigos sobrepostos, os mesmos já podem começar a ser tratados;</p> <p>13.1. Para a codificação em RS é mais indicado trabalhar blocos de texto;</p> <p>13.2. A codificação requer cuidados, principalmente para os casos em que se trabalha com blocos de textos e /ou entrevistas. Esta ressalva torna-se importante para que não seja empregada subjetividade dos pesquisadores nas descobertas e informações passadas pelos autores dos artigos selecionados;</p> <p>13.3. Os segmentos importantes do texto como conceitos, categorias e demais informações relacionadas com as dimensões consideradas devem ser rotulados e codificados;</p> <p>13.4. A codificação deve ser feita em um nível apropriado para as questões de pesquisa consideradas para a síntese temática.</p> <p>14. Verificar a codificação.</p> <p>14.1. A Verificação deve ser feita por outro pesquisador que não o que fez a codificação dos dados.</p> <p>14.2. Na verificação é importante garantir a consistência, aumentando a confiabilidade e permitindo aumento da credibilidade.</p> <p>14.3. Devem existir ligações claras entre o texto e os códigos identificados.</p> <p>14.4. É importante combinar com os demais envolvidos na síntese/verificação, os passos que serão adotados na execução das verificações, alinhando os demais critérios de verificação que venham a ser adotados.</p>
<p>Tradução de Código em Temas (Traduzir códigos em temas, subtemas e temas mais abrangentes)</p>	<p>15. Analisar se foi adquirida uma visão geral, inclusiva e compreensiva de todos os artigos;</p> <p>15.1. Com a divisão dos artigos em etapas iterativas de extração e codificação, consegue-se melhorar a visão geral, inclusiva e compreensiva de todos os artigos.</p> <p>15.2. Esta visão geral, inclusiva e compreensiva é importante para o sucesso deste passo, evitando retrabalhos.</p> <p>16. Traduzir os códigos em temas, considerando que:</p> <p>16.1. Códigos sobrepostos, que não tenham sido tratados até este passo, devem ser reduzidos;</p>

	<p>16.2. É importante separar os códigos que apresentem relações entre si utilizando um instrumento que melhore a visibilidade para esta análise;</p> <p>16.3. Análise dos códigos agrupados com a finalidade de tentar identificar um termo que possa abrangê-los (temas);</p> <p>16.4. Uma vez identificados, os temas devem ser verificados uns contra os outros;</p> <p>16.5. Os temas devem ser verificados com os dados dos artigos originais;</p> <p>16.6. Caso as extrações tenham sido transcrições de segmentos de texto dos artigos, a verificação mencionada para os temas pode ser feita em relação a estas extrações, não sendo imprescindível retornar aos artigos originais para esta atividade.</p> <p>16.7. A rastreabilidade deve ser mantida, entre os códigos gerados, temas e subtemas encontrados.</p> <p>17. Verificar a tradução dos códigos em temas.</p> <p>17.1. A Verificação deve ser feita por outro pesquisador que não o que fez a tradução dos códigos em temas.</p> <p>17.2. Na verificação deve-se tentar garantir que os temas são coerentes, consistentes e distintos.</p> <p>17.3. É importante combinar com os demais envolvidos na síntese/verificação, os passos que serão adotados na execução das verificações, alinhando os demais critérios de verificação que venham a ser adotados.</p>	
<p>Criação do Modelo de Alto Nível (Explorar relações entre temas e criar um modelo com os temas mais abrangentes)</p>	<p>18. Analisar, explorar e comparar as relações existentes entre os temas;</p> <p>19. Comparar as relações existentes entre os temas com as conclusões dos artigos;</p> <p>20. Criar um modelo para representar os temas e as relações existentes entre os temas;</p> <p>20.1. As saídas podem ser uma visão mais gráfica dos temas e suas relações (mapa temático), uma visão mais descritiva, uma taxonomia ou mesmo uma teoria.</p> <p>20.2. A rastreabilidade deve ser mantida de forma que pelo item do modelo de alto nível criado seja possível chegar ao código e, respectivamente, ao segmento de texto extraído do qual o mesmo foi originado.</p> <p>21. Descrever de forma clara as relações existentes entre os temas;</p> <p>21.1. O contexto em que as descobertas foram encontradas deve ser considerado. Portanto, as dimensões que serão codificadas auxiliam as comparações e conseqüentemente, auxiliam a criação dos modelos que representam os temas e suas relações.</p> <p>21.2. Revise o modelo de alto nível da última etapa, pegando cada ramo por vez, descrevendo o seu conteúdo com os resultados e o contexto das informações e criando temas de ordem superior.</p> <p>21.3. Identifique as ligações entre os temas de mais alto nível e as provas que fundamentam e contexto desta evidência.</p> <p>21.4. Explore as conexões com a teoria e pesquisas prévias, contextualize novamente, defina e ainda refine os temas de mais alto nível.</p> <p>22. Verificar a criação do modelo de alto nível.</p> <p>22.1. Esta deve ser feita por outro pesquisador ou por especialista no assunto.</p> <p>22.2. Na verificação deve-se tentar garantir que o modelo de alto nível gerado reflete as relações encontradas durante a realização do método.</p> <p>22.3. É importante combinar com os demais envolvidos na síntese/verificação, os passos que serão adotados na execução das verificações, alinhando os demais critérios de verificação que venham a ser adotados.</p> <p>23. Realizar a avaliação da confiabilidade da síntese.</p> <p>23.1. Para a realização da confiabilidade da síntese, preocupar-se com os conceitos e questões que compõem o <i>checklist</i> de avaliação da confiabilidade da síntese, sendo eles:</p>	
<p>Avaliação da Confiabilidade da Síntese (Avaliar a confiabilidade das interpretações que</p>	<p>Credibilidade</p>	<p>1. Os argumentos que conduzem às interpretações feitas na síntese temática foram explicitados?</p> <p>2. As suposições sobre uma abordagem específica para a síntese temática foram claramente explicadas?</p> <p>3. Existe um bom ajuste entre o que é reivindicado e o que</p>

antecederam a síntese temática)		a evidência mostra? 4. As questões de pesquisa foram respondidas com base na evidência da síntese temática? 5. A linguagem e os conceitos utilizados na síntese são consistentes?
	Confirmabilidade	6. Pesquisadores e especialistas concordam com a maneira que estes dados foram codificados e classificados?
	Confiança	7. Foi estabelecida e mantida a rastreabilidade das informações do artigo até a síntese?
	Transmissibilidade	8. Os resultados obtidos podem ser transferidos para outros contextos?

Com a execução desse método de síntese temática, tem-se como possíveis resultados: uma taxonomia (ferramenta adotada para classificar fenômenos com base em características comuns); um mapa temático (representação visual que permite a visualização de uma forma mais gráfica e ajuda a classificar os diferentes códigos em temas); ou uma teoria (conjunto de princípios relacionados com o assunto principal estudado na síntese).

Vale salientar que a escolha da técnica de análise de dados depende do pensamento e objetivo do pesquisador com a sua pesquisa.

2.7 CONSIDERAÇÕES FINAIS

Neste Capítulo foram apresentados os principais conceitos relacionados a Mapeamento Sistemático e Revisão Sistemática (elementos básicos, etapas, diferenças e outros), bem como, os tipos de pesquisa e suas técnicas de coleta e análise de dados com foco em pesquisas qualitativas.

A revisão realizada foi de grande importância para que houvesse uma imersão em processos sistemáticos e em análises qualitativas, base para a construção do método proposto nesta dissertação.

No capítulo seguinte é apresentado, de forma detalhada, o método para análise sistemática de grupos de pesquisa, objeto principal desta dissertação, juntamente com a descrição de cada atividade que o compõe.

Neste capítulo é apresentado o método para Análise Sistemática de Grupos de Pesquisa (MASG) proposto nesta dissertação, bem como, as ferramentas necessárias para a sua execução.

3 O MASG

O propósito principal desta dissertação é a apresentação de um método aqui chamado de MASG para realizar a análise sistemática de grupos de pesquisa e de seus pesquisadores. Este método tem como objetivo apoiar líderes de grupos e seus pesquisadores no direcionamento de suas pesquisas, bem como realizar uma caracterização e posicionamento do grupo em relação a uma área de interesse. A seguir esse método é descrito juntamente com as atividades que o compõem.

3.1 DESCRIÇÃO DO MÉTODO

A Figura 6 apresenta o MASG para realizar análise de um grupo de pesquisa em uma determinada área de interesse. Esse método pressupõe a existência e utiliza-se de informações de um Mapeamento Sistemático (MS) pré-existente e previamente escolhido. Estas informações incluem protocolo, taxonomia e mapas do MS. Elas são utilizadas como artefatos externos de entrada em algumas atividades do MASG com o objetivo de alinhar dados e possibilitar suas comparações.

O MASG tem como objetivo caracterizar o grupo e identificar as atuais estratégias de pesquisa adotadas por ele. Esta abordagem pode apoiar os líderes de

grupos de pesquisa e seus pesquisadores em suas tomadas de decisão quanto ao direcionamento das novas pesquisas do grupo.

O MASG é composto por um conjunto de sete atividades: (1) Identificar a fonte de dados; (2) Identificar os metadados e publicações do grupo; (3) Selecionar as publicações do grupo; (4) Extrair os dados das publicações do grupo; (5) Analisar a classificação das publicações do grupo; (6) Analisar e comparar os mapas; e (7) Entrevistar os pesquisadores do grupo.

A notação SPEM (*Software & System Process Engineering Metamodel*) foi utilizada para modelar o método. Essa notação foi escolhida devido ao seu destaque na literatura técnica relacionada à engenharia de processos e à sua forte ligação com a UML (*Unified Modeling Language*) (Münch *et al.*, 2012) o que dá mais clareza à representação.

A Figura 6 apresenta a sequência das atividades e os artefatos consumidos e produzidos em cada uma destas atividades. Os artefatos externos, provenientes do MS e o roteiro de entrevistas, são destacados na Figura em amarelo. As próximas subseções descrevem essas atividades.

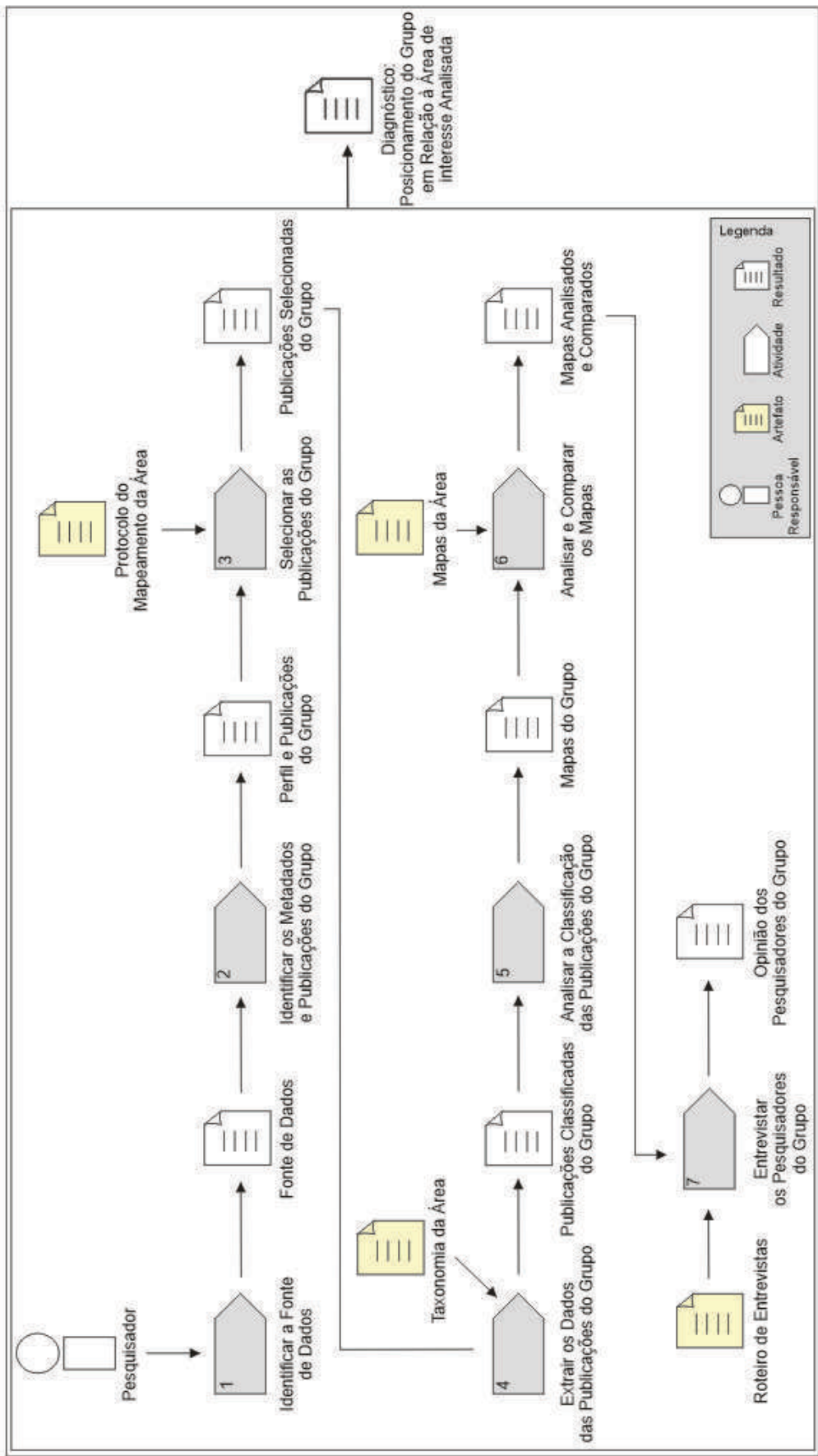


Figura 6. Método para Análise Sistemática de Grupos de Pesquisa (MASG)

3.1.1 Atividade 1 – Identificar a fonte de dados

A **Atividade 1 – Identificar a fonte de dados** – parte do objeto de análise, neste caso, o grupo, laboratório ou comunidade de pesquisa, juntamente com a área de interesse a ser analisada e o período de avaliação.

Com essas informações, o pesquisador deve identificar a(s) fonte(s) de dados que serão utilizadas para obter os dados das publicações dos pesquisadores do grupo. Essas fontes de dados devem conter não apenas artigos científicos publicados em eventos e periódicos, geralmente disponíveis em bibliotecas digitais tradicionais (por ex., *IEEE Xplore Digital Library*, *ACM Digital Library*, entre outras), mas também todas as publicações do grupo como: relatórios técnicos, dissertações, monografias, resumos, dentre outras publicações por pesquisador (por ex., Plataforma Lattes¹, DBLP² (*Digital Bibliography Library Project*), *Google Scholar*³, entre outras).

Sugere-se a utilização da Plataforma Lattes como fonte de dados, pois entende-se que os currículos (registro da formação e das produções acadêmicas de um pesquisador, além de sua experiência profissional), mantidos por este sistema, sejam confiáveis e devem estar atualizados devido à sua importância para a qualificação dos pesquisadores e instituições que atuam no Brasil.

Após a identificação da fonte de dados, deve-se também definir os métodos manuais ou automáticos que se utilizam de aplicativos computacionais para buscar os dados das publicações do grupo (por ex., nome, autor, coautores, ano de publicação, dentre outros) nas fontes de dados. Para essa busca é utilizado como entrada o nome dos pesquisadores do grupo.

¹ A plataforma Lattes é o principal sistema de informação científica, do Brasil, mantido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Disponível em: lattes.cnpq.br/

² A DBLP (*Digital Bibliography Library Project*) é uma biblioteca digital onde são encontrados dados mundiais relativos a publicações da área de ciência da computação. Disponível em: <http://dblp.uni-trier.de/>

³ Ferramenta de pesquisa onde são encontrados perfis de pesquisadores com suas produções científicas, coautores, índices de citações, etc. Disponível em: <https://scholar.google.com.br/>

Ferramentas como o ScriptLattes⁴ auxiliam nessa identificação, pois facilita a extração dos dados das publicações dos pesquisadores por meio dos seus currículos Lattes.

O ScriptLattes, além das características já apresentadas de identificação e por ser um sistema de código aberto, testado extensivamente em uma grande variedade de grupos (Mena-Chalco *et al.*, 2009), caracteriza-se como uma ferramenta que contempla o objetivo dessa atividade. Informações do grupo como: quantidade de publicações por tipo, grafo de colaboração dos pesquisadores, coautorias, dentre outros, são obtidos com a execução desse aplicativo.

Depois de identificar a fonte de dados e escolher o método de busca dos dados das publicações do grupo, deve ser realizada a identificação dos metadados e publicações do grupo. Esta atividade é descrita na próxima subseção.

3.1.2 Atividade 2 – Identificar os metadados e publicações do grupo

Na **Atividade 2 – Identificar os metadados e publicações do grupo** – é realizada a caracterização do grupo por meio da extração dos seus metadados e publicações. Entre os dados coletados estão à descrição do grupo, seus pesquisadores, linhas de pesquisa, quantidade de publicações, dentre outros.

A importância dessa atividade se dá pela apresentação sucinta e clara do grupo para a comunidade científica.

O Quadro 2 lista um conjunto de dados que podem ser coletados sobre o grupo e a descrição de cada um.

Quadro 2. Metadados do grupo

ID	Metadados	Descrições
1	Descrição do grupo	Permite identificar o grupo
2	Linhas de pesquisa do grupo	Permite identificar as áreas de interesse do grupo

⁴ <http://scriptlattes.sourceforge.net>

3	Nomes dos pesquisadores do grupo	Permite verificar a composição do grupo
4	Quantidade de publicações por pesquisador do grupo	Permite identificar quem são os pesquisadores que publicam no grupo
5	Quantidade de publicações do grupo por tipo de produção bibliográfica	Permite verificar os tipos de publicações bibliográficas do grupo
6	Quantidade de publicações do grupo por ano	Permite verificar a evolução das publicações do grupo
7	Grafo de colaboração dos pesquisadores do grupo	Permite visualizar a rede de relacionamento/colaboração dos pesquisadores do grupo

Esses dados são coletados por pesquisas manuais no site institucional do grupo ou similar (caso exista), por informações disponíveis em cadastros de grupos de pesquisa em órgãos competentes (ex. CNPq⁵) ou por obtenção dos dados diretamente com os líderes do grupo. Exemplos de dados são: quantidade de publicações por tipo, grafo de colaboração dos pesquisadores, dentre outros, os quais são extraídos por métodos de busca consistentes com as fontes definidas na **Atividade 1**.

Os dados adquiridos podem ser apresentados por meio de textos descritivos, como, por exemplo, o nome e as linhas de pesquisa do grupo; figuras utilizadas para apresentar informações adicionais como, por exemplo, o logotipo do grupo; e gráficos utilizados para apresentar de forma clara e visual informações quantitativas. Sugere-se a utilização de gráficos como o de barras e o de pizza para apresentar esses metadados do grupo.

Com o gráfico de barras pode-se verificar, por exemplo, os pesquisadores com o maior número de publicações no grupo e a quantidade de publicações do grupo por ano. A Figura 7 apresenta um exemplo de gráfico de barras que mostra a quantidade de publicações por pesquisador doutor do LES da UFBA no período de 2010-2015.

⁵ http://dgp.cnpq.br/dgp/faces/consulta/consulta_parametrizada.jsf

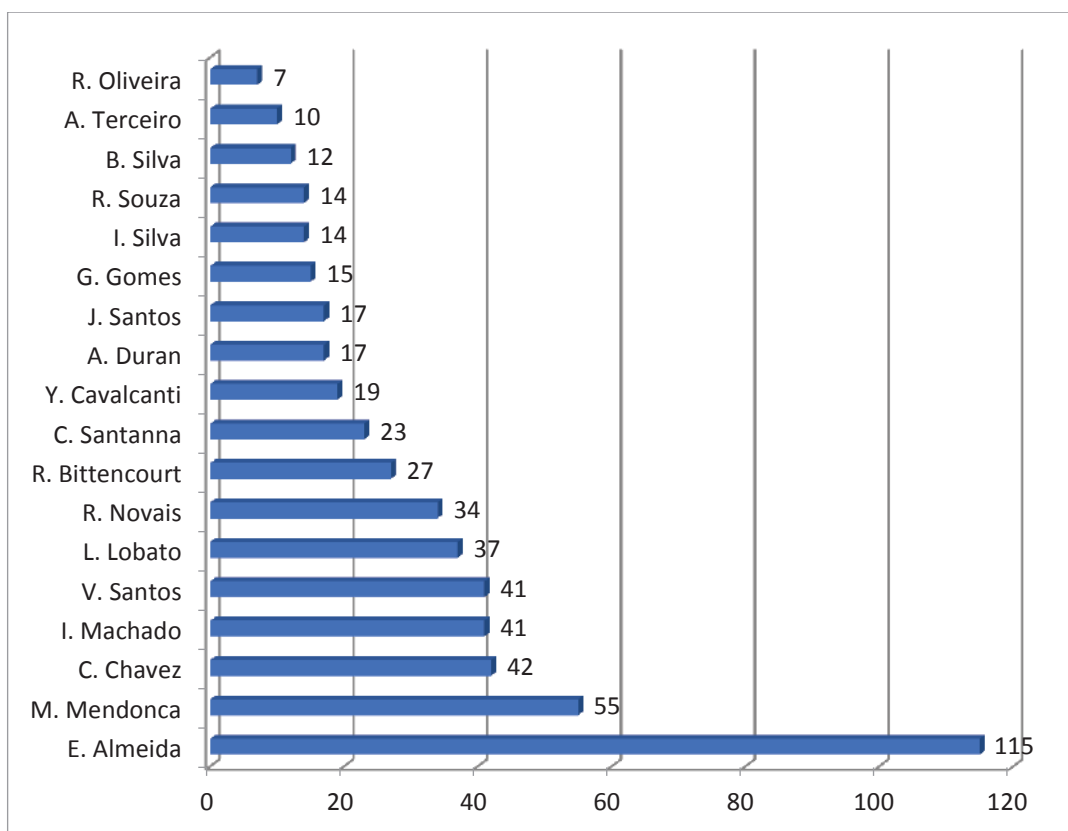


Figura 7. Exemplo de gráfico para mostrar as publicações por pesquisador

Com o gráfico de pizza pode-se representar as publicações do grupo por tipo, de forma que os pesquisadores possam verificar se precisam investir na produção de pesquisas de determinado tipo, com o objetivo de melhorar seus índices. A Figura 8 apresenta um exemplo de um gráfico de pizza das publicações bibliográficas do LES no período de 2010-2015 por tipo.

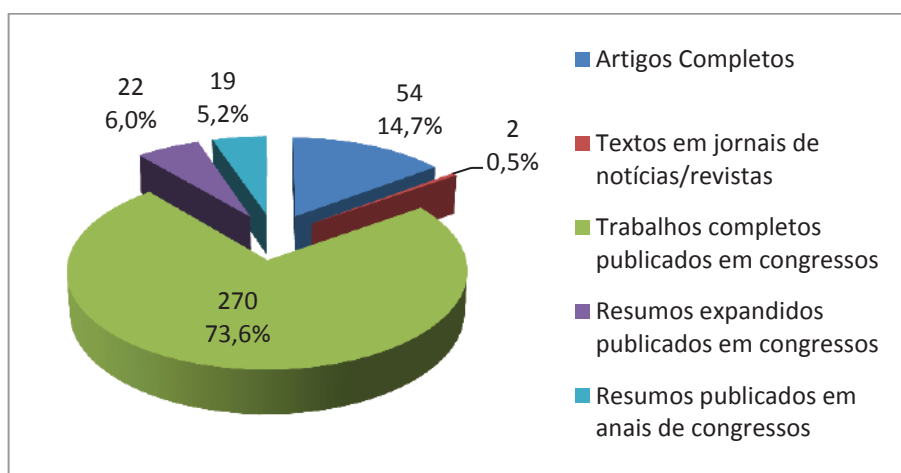


Figura 8. Exemplo de gráfico para mostrar a quantidade de publicações por tipo

Sugere-se também a utilização dos grafos de colaboração, pois servem para identificar como os pesquisadores do grupo se relacionam entre si. Essa rede de relacionamento/colaboração é importante para a criação de novas pesquisas e projetos, incentivam a criação do conhecimento e o processo de inovação resultantes do intercâmbio de informações e, sobretudo, da junção de competências de grupos que unem esforços na busca de metas comuns (Balancieri *et al.*, 2005). A Figura 9 apresenta o grafo de colaboração dos pesquisadores do LES do período de 2010-2015.

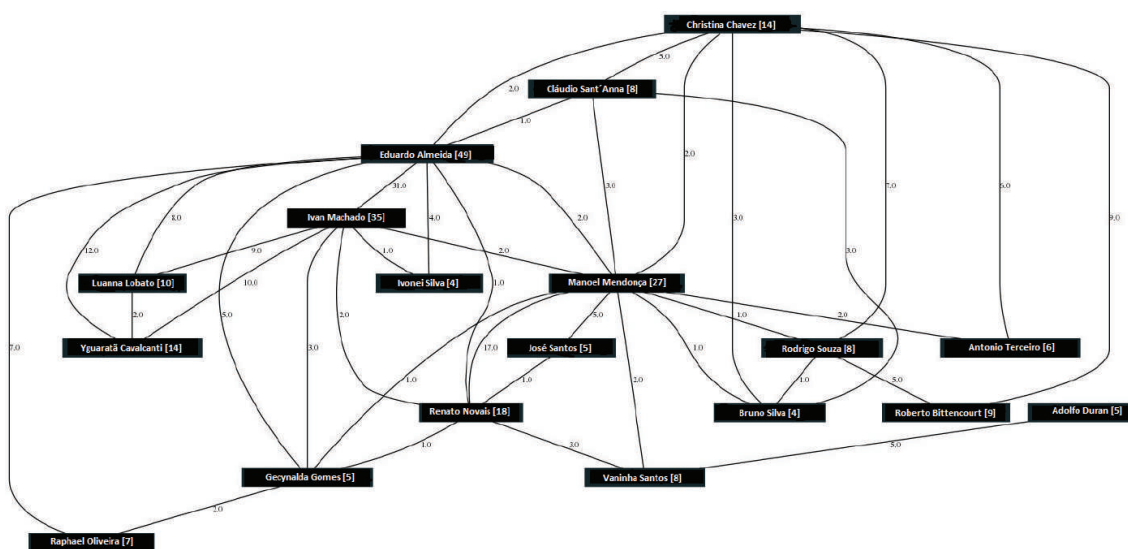


Figura 9. Exemplo de grafo de colaboração para mostrar a coautoria de publicações de pares de autores

Os dados extraídos das publicações do grupo podem ser listados em uma planilha eletrônica conforme Quadro 3 e disponibilizados como fonte de dados para consultas em *link* da internet.

É importante, caso o grupo não possua fontes conjuntas de publicações disponíveis para a comunidade, que os PDF's (*Portable Document Format da Adobe*) das publicações sejam reunidos em uma pasta e também disponibilizados online (ex. *Dropbox*, site do grupo, etc.), visto que, quando não existem essas fontes conjuntas de publicações, a maioria das publicações encontra-se descentralizadas, disponíveis em links dos próprios pesquisadores ou em bases de dados científicas (ex. *IEEE*, *Springer*, etc.). Essa centralização facilita a realização de novas pesquisas que utilizem essas publicações. Vale salientar que algumas publicações não podem ser divulgadas em

outros links, pois, possuem direito/licença exclusivos de algumas bases de dados científicas, neste caso devem ser divulgadas somente internamente para o grupo.

Quadro 3. Dados das publicações do grupo

Dado	Descrição
Tipo	Tipo de publicação. Ex. Artigo completo, resumo expandido, Trabalhos completos, etc.
Ano	Ano de publicação do trabalho.
ID	Número de identificação da publicação.
Título	Título da publicação.
Autores	Autores da publicação.
Fonte de publicação	Local onde foi publicado o trabalho.
Disponibilidade do PDF	Disponibilidade do PDF da publicação. Status: SIM ou NÃO.

Ao final desta atividade tem-se como resultado o perfil (caracterização) do grupo e suas publicações.

3.1.3 Atividade 3 – Selecionar as publicações do grupo

Para alcançar o objetivo de conhecer uma determinada área de interesse em um grupo, são requeridos direcionamentos e regras para selecionar as publicações relevantes, de forma a possibilitar a elaboração de conclusões importantes. Para obter esses direcionamentos e regras a **Atividade 3 – Selecionar as publicações do grupo** – consome o protocolo de um MS previamente escolhido, da área de interesse a ser analisada no grupo. Nele são discriminadas as questões de pesquisa e critérios de inclusão e exclusão utilizados para seleção das publicações, conforme objetivos de pesquisa.

A reutilização dos critérios de inclusão e exclusão e das questões de pesquisa possibilita o alinhamento da seleção das publicações do grupo com a seleção dos trabalhos do MS da área, o que viabiliza futuras comparações. Os critérios podem ser adaptados na busca de incluir vários tipos de publicações do grupo, como

dissertações de mestrado, teses de doutorado e trabalhos técnicos, visto que, a maioria dos mapeamentos sistemáticos excluem esses tipos de publicações, e consideram em suas análises somente artigos.

A seleção das publicações do grupo deve obedecer ao processo de triagem descrito na Figura 10, adaptado do estudo de Neto *et al.* (2011). Esse processo registra exclusões e inclusões de publicações no processo de cada filtro, o que possibilita o rastreamento dessas publicações. No Filtro 1 os pesquisadores selecionam as publicações pelo título. Para isso, aplicam-se os critérios de inclusão e exclusão reutilizados do protocolo do MS da área nos títulos das publicações. Caso haja dúvidas na seleção da publicação no Filtro 1, passa-se para o Filtro 2. Neste filtro a publicação deve ser recuperada no formato PDF, pois, o resumo e conclusão serão analisados apoiados nos mesmos critérios primeiramente utilizados nos títulos. Caso a dúvida ainda persista a publicação deve ser inteiramente lida para uma melhor avaliação (Filtro 3).

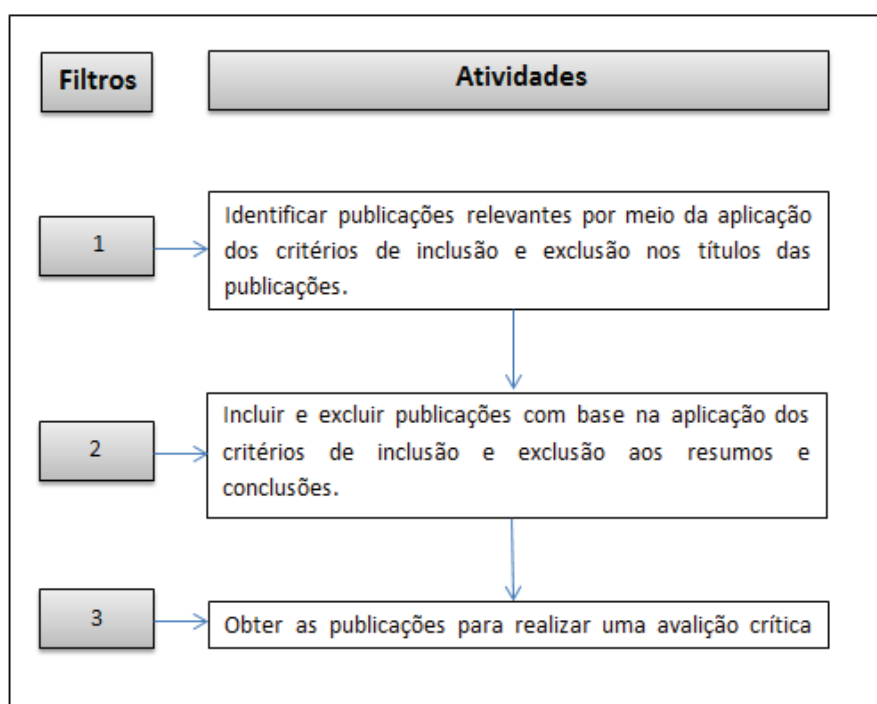


Figura 10. Processo de seleção das publicações (adaptado de Neto *et al.*, 2011)

Como resultados são obtidos as publicações selecionadas do grupo a serem consideradas na análise. Todos os PDF's das publicações selecionadas devem ser

obtidos, pois serão utilizados na atividade de extrair os dados das publicações do grupo, descrita na próxima subseção.

3.1.4 Atividade 4 – Extrair os dados das publicações do grupo

A **Atividade 4 – Extrair os dados das publicações do grupo** – tem como entrada a taxonomia da área, artefato externo extraído do MS da área. Essa taxonomia contempla uma estrutura classificatória que tem por finalidade servir de instrumento para organizar e recuperar informações (Bradley *et al.*, 2007). Ela contém uma lista estruturada de conceitos/termos de um domínio e deve estar disponível no MS da área de interesse escolhido. Caso essa taxonomia não esteja disponível, uma das alternativas é extraí-la por meio dos dados disponíveis no MS da área.

A definição dessa taxonomia é uma das atividades mais difíceis em mapeamentos sistemáticos. Dessa forma, sua reutilização para extração dos dados das publicações do grupo simplifica esse passo e garante mapas do grupo consistentes e compatíveis com os mapas criados pelo MS da área escolhido. Esses dados são importantes para as comparações que serão realizadas neste método.

Sugere-se que o objetivo da publicação seja um conceito pertencente à taxonomia, o que permite verificar se os objetivos das publicações do grupo estão alinhados aos objetivos dos trabalhos da área. Para isso, o modelo apresentado por Basili *et al.* (1999) pode ser utilizado para representar na taxonomia as categorias que apresentem esses objetivos.

Esse modelo tem normalmente quatro facetas: ponto de vista, objeto de análise, foco e propósito. Para extrair o objetivo da publicação basta somente três facetas: propósito, que é a principal intenção da publicação (ex. identificação, predição, compreensão (Basili *et al.*, 1999)); foco, que é o principal atributo de interesse do objeto de análise (ex. evolução do software, reuso de software, defeitos); e objeto de análise, que é a entidade de interesse analisada na publicação (ex. relatórios de *bugs*, código fonte, e-mails). Dividir o objetivo da publicação nessas três facetas é importante, pois facilita as posteriores comparações das mesmas.

Outro conceito que pode estar presente na taxonomia são os tipos de avaliações empíricas que foram utilizadas nas publicações. A extração dessa informação dá ao pesquisador a possibilidade de verificar o grau de confiança e qualidade das publicações analisadas. Esses métodos de avaliação podem ser classificados nos tipos estabelecidos por Wohlin *et al.* (2000) e Luders (2003), a depender de como o pesquisador descreveu o seu método de avaliação na publicação, a saber: estudo exploratório, estudo de caso, experimento controlado.

Os metadados das publicações (ex. autores, fontes de publicação, ano de publicação) também são informações importantes a serem extraídas. Segundo Novais *et al.* (2013) esses dados podem mostrar influências de pesquisadores que tem interesse na área de pesquisa.

O Quadro 7 do Capítulo 4, por exemplo, apresenta uma tabela de referência para extração dos dados das publicações que contém os conceitos sugeridos para essa atividade. Este é um artefato integrante do método, e é apresentado no próximo capítulo por questões didáticas.

O Quadro 4 apresenta um exemplo resumido de uma tabela de classificação da extração dos dados das publicações do LES que contempla a coleta dos dados descritos no Quadro 7, exceto os metadados. As categorias da classificação devem ser atualizadas conforme a área pesquisada. No exemplo, observa-se que a Publicação 1 (Farias *et al.*, 2015) tem o propósito de “compreensão” do “comportamento dos desenvolvedores” por meio do objeto de análise “e-mails” e “dados de *commits*” e utilizou o método de avaliação “estudo exploratório”. A Publicação 2 (Farias *et al.*, 2015a) tem o propósito “identificação” de “dívida técnica” por meio do objeto de análise “comentários” e utilizou o método de avaliação “estudo exploratório”. A Publicação 3 (Cavalcanti *et al.*, 2013) tem os propósitos de “identificação” e “compreensão” de “defeitos” por meio do objeto de análise “relatórios de *bugs*” e utilizou o método de avaliação “estudo de caso”.

Todos os significados dos dados devem ser definidos na busca de uma classificação organizada e lógica, o que facilita a recuperação das informações.

Uma taxonomia com as abordagens de coleta sugeridas para este método, na área temática de MSR, está disponível em endereço eletrônico⁶.

⁶ http://www.mestrado.web2405.uni5.net/Taxonomia_Atualizada.xlsx

Quadro 4. Exemplo resumido de tabela de classificação

	Identificação	Predição	Compreensão	Comportamento dos Desenvolvedores	Defeitos	Dívida Técnica	Dados de <i>Commits</i>	Comentários	E-mails	Relatórios de <i>Bugs</i>	Comparação com outras ferramentas	Estudo de Caso	Estudo Exploratório
Artigos	Propósito			Foco			Objeto de Análise				Método de avaliação		
Publicação 1			x	x			x		x				x
Publicação 2	x					x		x					x
Publicação 3	x		x		x					x		x	

Com o esquema de classificação definido, as publicações do grupo devem ser listadas em uma planilha eletrônica, com seus respectivos metadados (id, tipo, ano, título, resumo e autores) e classificação. Esta atividade tem o objetivo de obter informações relevantes que caracterizem o grupo na área de interesse escolhida. A planilha eletrônica deve ser disponibilizada em link da internet como fonte de dados para consultas e possíveis replicações do estudo.

Nessa atividade, critérios de confiabilidade da classificação devem ser adotados, entre eles temos as revisões por pares em que dois pesquisadores leem as publicações e as classificações resolvidas em uma etapa de consenso; caso os desentendimentos não sejam resolvidos por meio de discussões, um terceiro pesquisador é envolvido na classificação. Com esses critérios espera-se reduzir a probabilidade de erros.

Essa **Atividade 4** tem como resultado as publicações classificadas do grupo, utilizado na atividade de analisar a classificação das publicações do grupo, descrita na próxima subseção.

3.1.5 Atividade 5 – Analisar a classificação das publicações do grupo

A **Atividade 5 – Analisar a classificação das publicações do grupo** – tem como entrada de informação as publicações classificadas do grupo. Essa classificação é

analisada e apresentada por meio de recursos visuais como gráficos e figuras, em busca de tendências, *gaps* e informações relevantes do grupo em relação a área de interesse. Ela tem o objetivo de responder às questões de pesquisa definidas na **Atividade 3** de selecionar as publicações do grupo. Os gráficos e figuras devem ser apresentados da mesma forma que os apresentados no MS da área escolhido, o que possibilita as comparações que serão realizadas posteriormente.

Gráficos como os de bolhas são sugeridos para apresentar resultados que demandam uma análise mais complexa. Eles são bastante utilizados em mapeamentos sistemáticos (Farias *et al.*, 2016; Novais *et al.*, 2013; Münch *et al.*, 2012). A Figura 11 apresenta um exemplo de gráfico de bolhas que representa as facetas: propósito e método de avaliação das publicações do LES. Essa representação possibilita a verificação mais clara das frequências dessa combinação. Por exemplo, o gráfico mostra que, foram encontradas 16 publicações que utilizam o propósito “compreensão” e método de avaliação “estudo exploratório”.

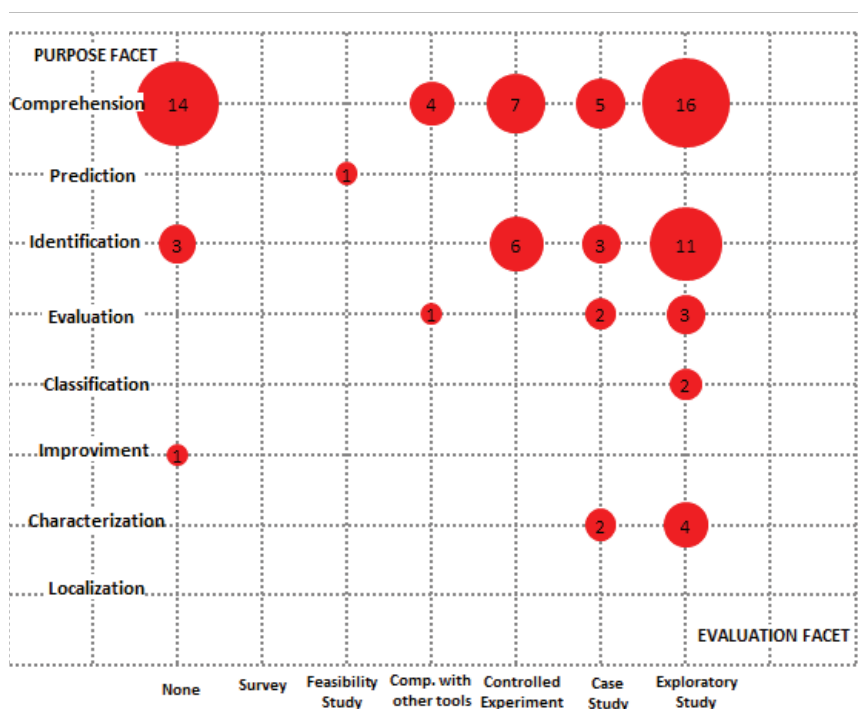


Figura 11. Exemplo de gráfico de bolhas para representar combinações de facetas

O resultado dessa **Atividade 5** são os mapas do grupo apresentados por meio de recursos visuais. Eles são semelhantes aos mapas produzidos durante o MS da

área. Eles contemplam as evidências encontradas durante as análises das classificações das publicações e são apresentados por meio de um resumo visual com sua classificação e categorização por resultados.

3.1.6 Atividade 6 – Analisar e comparar os mapas

A **Atividade 6 – Analisar e comparar os mapas** – tem como entradas os mapas do grupo e os mapas da área. Os mapas da área são artefatos externos adquiridos no MS da área escolhido e contém a análise dos seus resultados com a frequência dos trabalhos por categoria dentro de um determinado esquema de classificação. Esses mapas possuem informações relevantes da área analisada. Os mapas do grupo contêm as mesmas informações do grupo relacionadas à mesma área de interesse. Essa semelhança só é possível por causa da reutilização da taxonomia do MS da área escolhido para classificar as publicações do grupo, realizada na **Atividade 4**.

Nessa atividade os mapas do grupo devem ser analisados de forma a compará-los com os mapas da área de interesse. Essa comparação sistemática dos mapas:

- Ajuda os pesquisadores a entenderem como estão em relação a uma área de interesse;
- Permite comparar pesquisadores diferentes na mesma área de interesse de forma sistemática;
- Facilita a identificação de pontos de sinergia e complementariedade entre os pesquisadores;
- Permite identificar gaps da área que não estão sendo trabalhados pelo grupo.

Todas as comparações são feitas de forma sistemática e permitem que as publicações de uma área de interesse de um grupo de pesquisa sejam comparadas com os trabalhos da comunidade acadêmica da mesma área. A utilização do MASG permite também que essas comparações sejam feitas com outros grupos de pesquisa que apliquem esse mesmo método para a mesma área de interesse. Sem o MASG os

pesquisadores podem até verificar como eles estão em relação a uma área de interesse, mas não vão poder fazer comparações com outros pesquisadores, pois isso só pode ser feito por meio de um processo sistemático.

Para as comparações dos mapas, recursos visuais devem ser aplicados em busca de uma melhor visualização de semelhanças e diferenças entre eles.

Para essa atividade podem ser utilizados gráficos de colunas para comparar dados numéricos. A Figura 12 apresenta um gráfico de colunas com a comparação dos propósitos das publicações do LES na área relacionada à Mineração de Repositório de Software (MSR) e dos resultados do MS da mesma área no período de 2010-2014. Com este gráfico observam-se com clareza os propósitos que são mais utilizados e suas proporções.

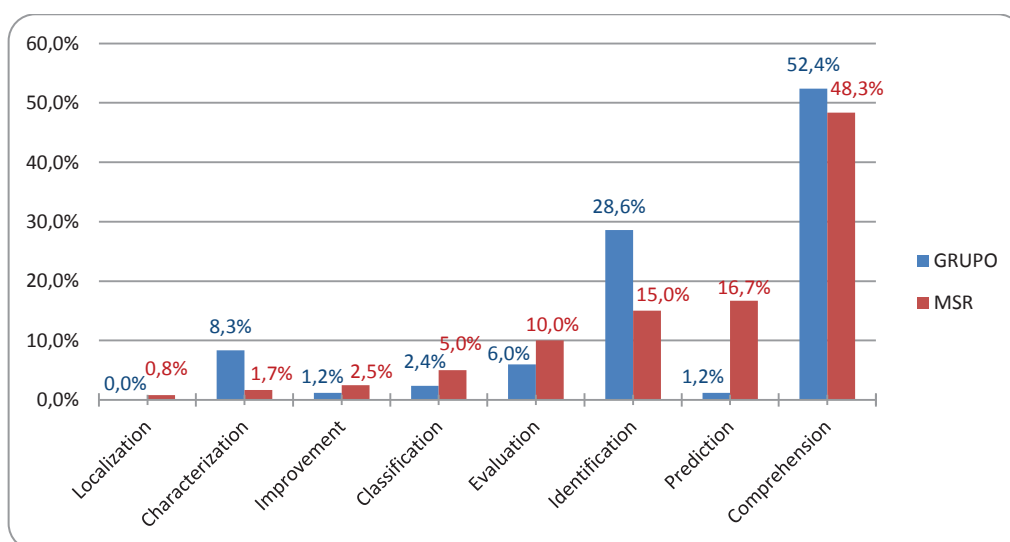


Figura 12. Exemplo de gráfico de colunas para comparar dados numéricos

Os gráficos de bolhas também devem ser utilizados nessa atividade. Sugere-se que esses gráficos sejam gerados igualmente para um determinado resultado da área e do grupo e sejam sobrepostos um ao outro, com as bolhas em cores diferentes. Essa sobreposição permite a visualização mais clara das diferenças entre os resultados. Pelo nosso conhecimento, esse tipo de comparação nunca foi abordada em estudos científicos. Dessa forma, este tipo de comparação, caracteriza-se como uma contribuição desta dissertação.

A Figura 13 apresenta um exemplo de sobreposição dos resultados de 2 gráficos de bolhas. Essa sobreposição permite comparar a frequência das combinações entre propósito e método de avaliação. Com bolhas vermelhas apresenta-se o gráfico de bolhas resultado das publicações do LES e com bolhas sem cor e contorno preto o resultado dos trabalhos da área. A variação de cores deixa as diferenças entre os resultados mais fáceis de serem visualizadas.

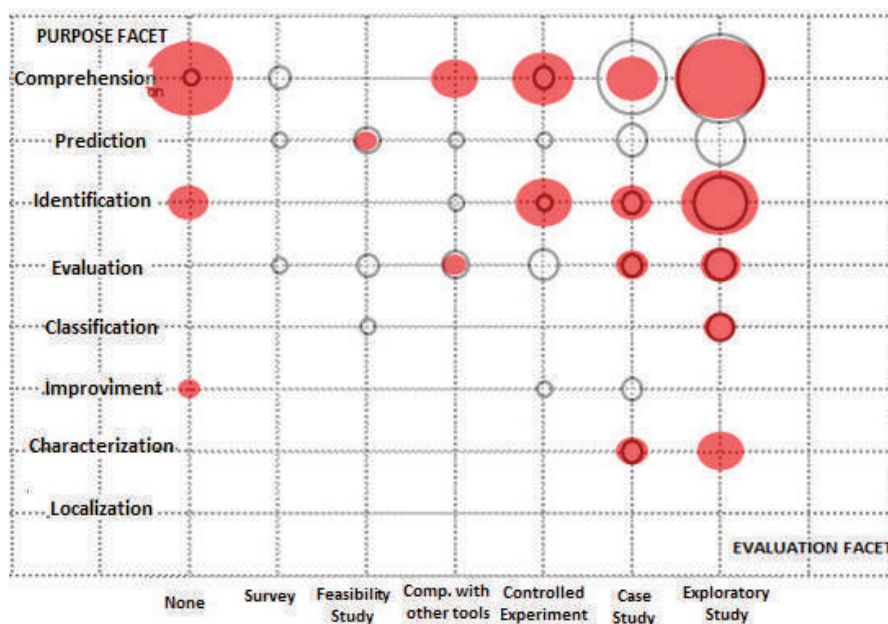


Figura 13. Exemplo de sobreposição de gráficos de bolhas para visualizar diferenças entre resultados

Após as comparações feitas por meio dos recursos visuais, os resultados das análises comparativas devem ser descritos em textos explicativos.

O resultado dessa atividade são os mapas analisados e comparados, utilizados como fonte de informação para a atividade de entrevistar os pesquisadores do grupo apresentada na próxima subseção.

3.1.7 Atividade 7 – Entrevistar os pesquisadores do grupo

Essa **Atividade 7 - Entrevistar os pesquisadores do grupo** – tem como objetivo extrair informações diretamente dos pesquisadores para alinhar e concluir a pesquisa. Ela utiliza o roteiro de entrevistas descrito no Quadro 5 do Capítulo 4, que tem o objetivo de direcionar a coleta de dados dos pesquisadores em relação aos resultados produzidos durante a execução do método. O roteiro é item integrante do conjunto de artefatos desenvolvidos para ser utilizado no método.

No roteiro de entrevistas as principais comparações dos mapas da área e do grupo e também do pesquisador de forma individual são apresentadas ao pesquisador, para melhorar a sua visão e análise na resposta aos questionários das entrevistas, e obter sua posição de forma clara em relação aos dados apresentados. O questionário pode conter, mas não precisa se limitar, às perguntas apresentadas no Quadro 5 do Capítulo 4.

Segundo Lakatos e Maroni (2010) as dificuldades e problemas que acontecem durante uma pesquisa que envolve coleta de dados qualitativos, nem sempre podem ser previstas. A aplicação de pré-testes, ou seja, aplicação do questionário, em uma versão preliminar, numa amostra de indivíduos deve ser adotada. Esses pré-testes melhoram a qualidade do questionário definitivo e diminuem a possibilidade de erros com a sua reformulação.

O planejamento prévio é essencial para realização desta coleta de dados. A rigorosa aplicação desse roteiro e questionário é de fundamental importância para o bom andamento da pesquisa.

As entrevistas tendem a esclarecer evidências que não estão claras nos resultados prévios e levantar novas informações importantes para a conclusão do diagnóstico. Elas devem ser realizadas presencialmente com uma amostra dos pesquisadores do grupo que apresentam publicações na área de interesse escolhida.

De posse de todos os resultados obtidos com a execução do método e realizadas as suas análises, é apresentado o diagnóstico ou posicionamento do grupo em relação a área de interesse analisada.

Esse diagnóstico contempla as atividades de pesquisa do grupo, sua caracterização e estratégias de pesquisa que vêm sendo adotadas pelo mesmo. Esses

dados devem ser analisados de forma quantitativa (ex.: para transformar os dados em números, o que possibilita análises numéricas) e/ou qualitativa (ex.: para entendimento da percepção dos textos das entrevistas). A síntese temática é um dos métodos para análise de dados qualitativos que pode ser utilizado.

3.2 CONSIDERAÇÕES FINAIS

Neste Capítulo foi apresentado o método proposto por este trabalho, o MASG – Método de Análise Sistemática de Grupos de Pesquisa, juntamente com a descrição de todas as atividades que o compõe. Este método é a principal contribuição deste trabalho.

O MASG, além de caracterizar um grupo de pesquisa em relação a uma área de interesse, possibilita que líderes de grupos de pesquisa possam (1) tomar decisões quanto à direção das novas pesquisas a serem realizadas pelo grupo e (2) identificar as estratégias de pesquisa que estão sendo adotadas pelo grupo. Além disto, a MASG permite que os pesquisadores do grupo entendam como estão atuando em relação a área, tanto em comparação com os outros membros do grupo, quanto em comparação com a área como um todo.

No próximo capítulo serão apresentados os artefatos produzidos e/ou utilizados durante o desenvolvimento deste método e que podem ser reutilizados para estudos futuros com outros grupos de pesquisa (ex. roteiro de entrevista, tabela de referência para extração dos dados às publicações, dentro outros).

Eles foram referenciados durante a apresentação de cada atividade do MASG.

Neste capítulo são descritos os artefatos utilizados e/ou construídos durante a utilização do método proposto nesta dissertação e que podem ser reutilizados em outros estudos.

4 ARTEFATOS UTILIZADOS NO MÉTODO

4.1 MAPEAMENTO SISTEMÁTICO DA ÁREA

Esta seção apresenta o mapeamento sistemático (MS) da área que foi escolhido para coletar alguns artefatos de entrada necessários para as atividades que serão realizadas durante a execução do Método para Análise de Grupos de Pesquisa (MASG).

Para avaliação do MASG, utilizou-se alguns artefatos produzidos no MS apresentado no estudo de Farias *et al.*, (2016) (Anexo 1), o qual teve a participação do autor desta dissertação, pois apresenta mapas atuais da área de Mineração de Repositório de Software (MSR) e dispõe dos seguintes artefatos: protocolo, taxonomia e mapas, o que supri as necessidades e satisfaz as exigências do método proposto.

Esse mapeamento sistemático teve como objetivo avaliar a área de MSR pela identificação do objetivo dos estudos através do seu propósito, foco e objeto de análise. As fontes de dados, os métodos de avaliação, as ferramentas e a evolução da área também foram analisados.

Ele utilizou uma busca manual na *International Conference on Mining Software Repositories* (MSRConf)⁷. Foram avaliados 107 artigos completos contidos na MSRConf dos anos de 2010 a 2014.

A taxonomia utilizada no processo de classificação dos dados contendo conceitos/termos deste MS está apoiada pelo artigo de Kagdi *et al.* (2007) que traz em seu estudo uma taxonomia expressiva para a área de MSR, pelo estudo de Novais *et al.* (2013) e por conversas com especialistas. Essa junção possibilitou a apresentação de um conjunto de categorias que são representativas da população estudada e está disponível em endereço eletrônico⁸.

No protocolo deste MS apresenta-se a questão geral e as específicas de pesquisa e os critérios de inclusão e exclusão utilizados para selecionar os trabalhos relevantes para a pesquisa, os mesmos são listados abaixo:

Questão Geral de Pesquisa (QG): “*Quais são as metas atuais abordadas por estudos recentes em MSR?*”.

Questões Específicas (QE) de pesquisa:

QE1. *Quais são os principais propósitos, foco e objeto de análise, identificados nos estudos?*

QE2. *Quais são os tipos de fontes de dados suportados pela Mineração de Repositório Software e como elas estão evoluindo?*

QE3. *Quais são as avaliações empíricas realizadas?*

Critério de inclusão: artigos completos pertencentes ao domínio de Engenharia de Software, relacionados à área de MSR e publicados na MSRConf.

Critérios de exclusão: (i) estudos que não abordam Mineração de Repositório de Software no domínio de Engenharia de Software e (ii) resumos de tutoriais, artigos curtos, *surveys*, estudos experimentais secundários, *challenge* e *showcase*.

Vale salientar que os resultados do MS de Farias *et al.* (2016) estão apresentados no [Anexo 1](#). Outras figuras e resultados apresentados durante este trabalho que não estão presente no MS foram trazidos de documentos de análise produzidos pelo

⁷ msrconf.org/

⁸ <http://goo.gl/hk67cc>

autor desta dissertação durante a execução do mapeamento, com o objetivo de mais à frente, compará-los com os resultados ou mapas do grupo a serem analisados.

4.2 ROTEIRO DE ENTREVISTAS

Esta seção apresenta na Figura 13 o roteiro de entrevistas adotado na execução do Método para Análise de Grupos de Pesquisa (MASG) apresentado nesta dissertação.

Quadro 5. Roteiro de entrevistas

ROTEIRO DE ENTREVISTA
DADOS DE IDENTIFICAÇÃO
<p>Nome do Entrevistado: _____</p> <p>Nome do Entrevistador: _____</p> <p>Objetivo da Entrevista Verificar a visão do pesquisador em relação à área, seu entendimento sobre as análises e capturar sua visão sobre o seu papel e o papel do grupo de pesquisa na área de conhecimento escolhida.</p>
METADADOS DO GRUPO, LABORATÓRIO OU COMUNIDADE DE PESQUISA
<p>Nesta seção os dados visuais dos metadados do grupo, produzidos na execução da Atividade 2 da Figura 6, devem ser apresentados ao entrevistado para que ele possa se posicionar em relação aos mesmos. Durante a apresentação o entrevistado pode esclarecer possíveis dúvidas quanto aos dados. Abaixo dados que devem ser apresentados:</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Nome do grupo, laboratório ou comunidade de pesquisa analisada com logotipo (caso possua) <input checked="" type="checkbox"/> Quantidade de publicações analisadas por pesquisador do grupo <input checked="" type="checkbox"/> Quantidade de publicações analisadas do grupo por tipo <input checked="" type="checkbox"/> Quantidade de publicações analisadas do grupo por ano <input checked="" type="checkbox"/> Grafo de colaboração dos pesquisadores do grupo
RESULTADO DAS COMPARAÇÕES DOS MAPAS
<p>Nesta seção os dados visuais das comparações realizadas entre os mapas da área de conhecimento escolhida e os mapas do grupo, produzidos na execução da Atividade 6 da Figura 6, devem ser apresentados ao entrevistado para que ele possa se posicionar em relação aos mesmos. Durante a apresentação o entrevistado pode esclarecer possíveis dúvidas quanto aos dados. Abaixo dados que devem ser apresentados:</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Comparação dos propósitos dos estudos <input checked="" type="checkbox"/> Comparação dos focos dos estudos <input checked="" type="checkbox"/> Comparação dos objetos de análise dos estudos <input checked="" type="checkbox"/> Comparação dos métodos de avaliação utilizados pelos estudos <p>Demais dados visuais devem ser apresentados conforme escolha da área de conhecimento escolhida e extração realizada.</p>

DADOS DO PESQUISADOR EM RELAÇÃO AOS MAPAS

Nesta seção os dados visuais do entrevistado devem ser apresentados, comparando-os aos resultados das comparações dos mapas da área de conhecimento escolhida e dos mapas do grupo.

Durante a apresentação o entrevistado pode esclarecer possíveis dúvidas quanto aos dados.

Abaixo dados que devem ser apresentados:

- ✗ Nomes e quantidade de publicações analisadas do pesquisador na área de conhecimento
- ✗ Comparação dos propósitos dos estudos
- ✗ Comparação dos focos dos estudos
- ✗ Comparação dos objetos de análise dos estudos
- ✗ Comparação dos métodos de avaliação utilizados pelos estudos

Demais dados visuais devem ser apresentados conforme escolha da área de conhecimento escolhida e extração realizada.

QUESTIONÁRIO

Primeira Etapa

Objetivo

Capturar a visão do pesquisador em relação à área analisada.

PERGUNTAS

1. Esta área de conhecimento é a sua principal área de pesquisa ou somente uma área de apoio? Explique.

Com essa pergunta pretende-se identificar o grau de importância da área de conhecimento analisada para os pesquisadores.

2. Qual o seu principal objetivo e motivação para elaboração de estudos nesta área? Explique.

Com essa pergunta pretende-se identificar porque os pesquisadores realizam pesquisas na área de conhecimento analisada.

3. As comparações dos mapas lhe permitem entender melhor a área de conhecimento e ter novas visões e direcionamentos para os seus estudos? Explique.

Com essa pergunta pretende-se identificar a importância das comparações dos mapas para os pesquisadores, bem como saber se essas comparações influenciam no direcionamento das suas atividades de pesquisa.

QUESTIONÁRIO

Primeira Etapa

Objetivo

Capturar os objetivos e estratégias utilizadas pelo grupo.

PERGUNTA

1. Abaixo apresentamos uma lista de classificação de objetivos. Gostaríamos que você marcasse os objetivos que você utiliza e relatasse quais estratégias estão sendo utilizadas para alcançar esses objetivos.

	Objetivo	Estratégia
	Captação de Recursos Financeiros	
	Ampliação do Grupo	
	Melhoria dos Recursos Humanos	
	Desenvolvimento de Parcerias com outros Grupos	
	Desenvolvimento de Parcerias com o setor Produtivo	
	Identificação de Novas Oportunidades de Pesquisa	
	Melhoria de Infraestrutura de Pesquisa	
	Nucleação de Novos Grupos	
	Outros. Identificar. _____	

4.3 TERMO DE CONSENTIMENTO

Esta seção apresenta o modelo do termo de consentimento adotado na execução do Método para Análise de Grupos de pesquisa e Individuais (MASG).

Quadro 6. Modelo do termo de consentimento para entrevistas

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	
<p>Prezado(a) participante:</p> <p>Sou _____ do curso de _____ da _____. Estou realizando uma pesquisa sob supervisão do Prof. _____, cujo objetivo é _____.</p> <p>Sua participação envolve uma entrevista, que será gravada se assim você permitir.</p> <p>A participação nesse estudo é voluntária e se você decidir não participar ou quiser desistir de continuar em qualquer momento, tem absoluta liberdade de fazê-lo.</p> <p>Na publicação dos resultados desta pesquisa, sua identidade será mantida no mais rigoroso sigilo. Serão omitidas todas as informações que permitam identificá-lo.</p> <p>Mesmo não tendo benefícios diretos em participar, indiretamente você contribuirá para a compreensão do fenômeno estudado e para a produção de conhecimento científico.</p> <p>Quaisquer dúvidas relativas à pesquisa poderão ser esclarecidas pelo responsável _____, fone _____ ou pelo estudante fone _____.</p> <p style="text-align: right;">Atenciosamente,</p> <p style="text-align: center;">_____ Pesquisador Matrícula: xxxxx</p> <p style="text-align: right;">Cidade-UF, ___ de _____ de _____</p> <p style="text-align: center;">_____ Orientador</p> <p>Consinto em participar deste estudo e declaro ter recebido uma cópia deste termo de consentimento.</p> <p style="text-align: center;">_____ Entrevistado</p> <p style="text-align: right;">Cidade-UF, ___ de _____ de _____</p>	

4.4 TABELA DE REFERÊNCIA PARA EXTRAÇÃO DOS DADOS DAS PUBLICAÇÕES

Essa seção apresenta no Quadro 7 uma tabela de referência para extração dos dados das publicações que contém os conceitos sugeridos para execução da **Atividade 4** do MASG.

Quadro 7. Tabela de referência para extração dos dados das publicações

Dados	Descrição
Título, Ano e Autores	Estes dados são chamados de metadados dos estudos. As extrações destes dados possibilitam identificar autores e ano de publicação dos estudos. Segundo Novais <i>et al.</i> (2013) esses dados podem demonstrar influências de pesquisadores e instituições que mostram interesse na área de pesquisa.
Fonte de Dados	Fontes de dados ou repositórios utilizados pelos pesquisadores para extrair as informações de interesse do estudo.
Propósito	Intenção principal das publicações, ou seja, o seu principal objetivo (Basili <i>et al.</i> , 1999). Exemplos: Avaliação, predição, compreensão, dentre outros.
Foco	Principal atributo de interesse do objeto de análise (Basili <i>et al.</i> , 1999). Exemplos. <i>Smells</i> , comportamento de desenvolvedores, etc.
Objeto de Análise	Entidade de interesse analisada no estudo (Basili <i>et al.</i> , 1999). Exemplos: código fonte, <i>commits</i> , e-mails, etc.
Método de Avaliação	Tipos de avaliações/experimentações feitas nos estudos. Segundo Basili e Weiss (1984) este dado é importante para verificar a confiança dos resultados e repetição do estudo no mesmo ou em diferentes ambientes (estudo de caso, experimentos, entrevistas para obtenção de dados). Exemplos: estudo de caso, experimentos controlados, etc.

4.5 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os artefatos utilizados do MS da área de MSR (taxonomia, protocolo e mapas) e/ou produzidos (roteiro de entrevista, termo de consentimento, tabela de referência para extração dos dados das publicações) durante o desenvolvimento do método proposto. O MS não é uma contribuição desta dissertação. A autora apenas colaborou na sua execução. Os artefatos produzidos são uma contribuição deste trabalho, parte integrante do método apresentado e podem ser reutilizados em outros estudos.

No próximo capítulo será apresentado um estudo de caso realizado no Laboratório de Engenharia de Software da UFBA (LES) com o objetivo de avaliar o método proposto nesta dissertação. Nesse estudo foram utilizados todos os artefatos descritos neste capítulo.

Neste capítulo é apresentado o estudo de caso realizado para mostrar a aplicabilidade prática do Método para Análise Sistemática de Grupos de Pesquisa (MASG) proposto no capítulo 3 desta dissertação, bem como, todos os resultados obtidos durante o seu processo.

5 ESTUDO DE CASO

Com o objetivo de avaliar o Método de Análise Sistemática de Grupos de Pesquisa (MASG), que possibilita a caracterização de um grupo de pesquisa em uma determinada área de interesse, bem como a identificação das estratégias de pesquisa utilizadas por ele, foi realizado um estudo de caso no Laboratório de Engenharia de Software (LES) da Universidade Federal da Bahia (UFBA).

Nesse estudo de caso, publicações científicas dos LES entre os anos de 2010-2015 foram avaliadas. Diante a disponibilidade dessas publicações, a área de Mineração de Repositório de Software (MSR) foi escolhida como área temática para este estudo.

Para a análise do LES foram utilizadas publicações até o ano de 2015, portanto, um ano a mais com relação às publicações analisadas da área (2010-2014). Isto foi feito, pois o MS da área de MSR (Anexo 1) já estava finalizado e, ao contrário da área, a análise do LES ainda seria desenvolvida, e se beneficiaria de uma análise mais abrangente para produção de resultados mais significativos.

Nesse contexto, na próxima seção, com a utilização do cenário descrito acima, são apresentados os resultados obtidos com a aplicação do MASG.

5.1 ANÁLISES DO GRUPO DE PESQUISA

O método proposto neste trabalho (Figura 6) foi utilizado neste estudo de caso e todas as atividades do processo foram realizadas. Os resultados de cada atividade do método são apresentados nas subseções abaixo.

5.1.1 Atividade 1 – Identificar a fonte de dados

Na **Atividade 1 – Identificar a fonte de dados** – para identificação das publicações desenvolvidas pelo LES, a Plataforma Lattes, sugerida pelo método proposto (Figura 6), foi escolhida para realizar esta análise. Essa fonte de dados é composta pelos currículos lattes dos pesquisadores e estudantes do país e se tornou um padrão nos registros dos resultados científicos da comunidade acadêmica nacional.

Essa plataforma é mantida pelo Conselho Nacional de Pesquisa e Desenvolvimento (CNPQ⁹). Os seus currículos são adotados pela maioria das instituições de fomento, universidades e institutos de pesquisa do país e se tornaram elemento indispensável e compulsório à análise de mérito e competência dos pleitos de financiamentos nacionais na área de ciência e tecnologia. Diante este contexto entende-se que a Plataforma Lattes é uma fonte confiável e atualizada de informações acadêmicas.

Como método automático de identificação das publicações produzidas pelos pesquisadores do LES, a aplicação computacional ScriptLattes, sugerida pelo MASG, foi escolhida devido a sua capacidade de agrupar as publicações informadas pelos pesquisadores em seus currículos lattes.

⁹ www.cnpq.br/

5.1.2 Atividade 2 – Identificar os metadados e publicações do grupo

Na **Atividade 2 – Identificar os metadados e publicações do grupo** – pesquisas manuais foram realizadas no site do CNPq¹⁰ e no site do LES¹¹ para coletar os metadados do laboratório e com isso identificá-lo.

Nos sites mencionados foram extraídos os seguintes dados:

O LES, por meio dos seus grupos, investiga subáreas da Engenharia de Software (ES), como, visualização de software, linhas de produto de software, evolução de software, qualidade de software, dentre outras, com o objetivo de estudar disciplinas de ES, assim como as áreas que impactam em como são desenvolvidos, mantidos e gerenciados os softwares.

Pertencem ao LES, cinco grupos de pesquisa (Figura 14), são eles: CEMantika – *Context and Ubiquitous Systems* (2009), aSide – *Software Design and Evolution* (2003), MEFES – *Formal Methods in Software Engineering* (2009), RISE – *Reuse software Engineering* (2011) e SoftVis – *Software Visualization Groups* (2010).



Figura 14. Grupos que compõe o LES

¹⁰ http://dgp.cnpq.br/dgp/faces/consulta/consulta_parametrizada.jsf

¹¹ <http://wiki.dcc.ufba.br/LES/>

Alguns desses grupos são translaboratoriais, ou seja, possuem pesquisadores de outros grupos (ex. RISE), e outros até mesmo interinstitucionais, ou seja, trabalha em grupo, parceria ou colaboração com uma ou mais instituições que se identificam pela possibilidade de desenvolver projetos ou objetivos comuns (ex. SoftVis).

Cada grupo do LES possui linhas de pesquisas ou interesses distintos, as mesmas estão listadas por grupo no Quadro 8.

Quadro 8. Linhas de pesquisa por grupo do LES

Grupo de Pesquisa	Linhas de Pesquisa ou Interesses
CEMantika	Contexto Computacional, Modelos matemáticos e metamodelos para sistemas sensíveis ao contexto, Sistemas de Transporte Inteligentes Ubíquos, Sistemas de recomendação móveis usando informações georreferenciadas, RFID, Internet das Coisas, Redes Sociais e Crowdsourcing e Dispositivos Moveis e Mineração de texto.
aSide	Arquitetura de software, Evolução de software, Qualidade de Software, Desenvolvimento de Software Orientado a Aspectos e Educação em Engenharia de Software.
MEFES	Métodos Formais em Engenharia de Software
RISE	Ferramentas CASE e ambientes, Componente de software, Engenharia de Software Empírica, Arquitetura de software e Linhas de Produto de Software (SPL).
SoftVis	Visualização de software, Engenharia de Software Experimental, Evolução de software e Compreensão de software.

Para identificação dos pesquisadores foram realizadas reuniões com o líder do LES e do grupo SoftVis. Nessas reuniões foram identificados os pesquisadores a serem analisados nesta pesquisa. Utilizou-se o seguinte critério de inclusão: pesquisadores de cada grupo que trabalham e/ou trabalharam no LES entre o período de 2010-2015 na área de ES da UFBA e que possuem doutorado.

A quantidade de pesquisadores obedece à descrição seguinte: o grupo CEMantika possui 1 pesquisador que é o líder do grupo, o grupo aSide possui 6 pesquisadores sendo 2 líderes do grupo, o grupo MEFES com 1 pesquisador que é o líder do grupo, o grupo RISE com 7 pesquisadores sendo 2 líderes do grupo e o grupo SoftVis com 3 pesquisadores sendo 1 líder do grupo. Todos foram incluídos no estudo,

o que totaliza 18 pesquisadores. A lista dos pesquisadores analisados está disponível no Apêndice A.

Com os nomes dos pesquisadores que compõem o LES e o período de análise, a ferramenta ScriptLattes foi executada.

A Figura 15 apresenta os resultados das produções do LES identificadas pelo ScriptLattes.

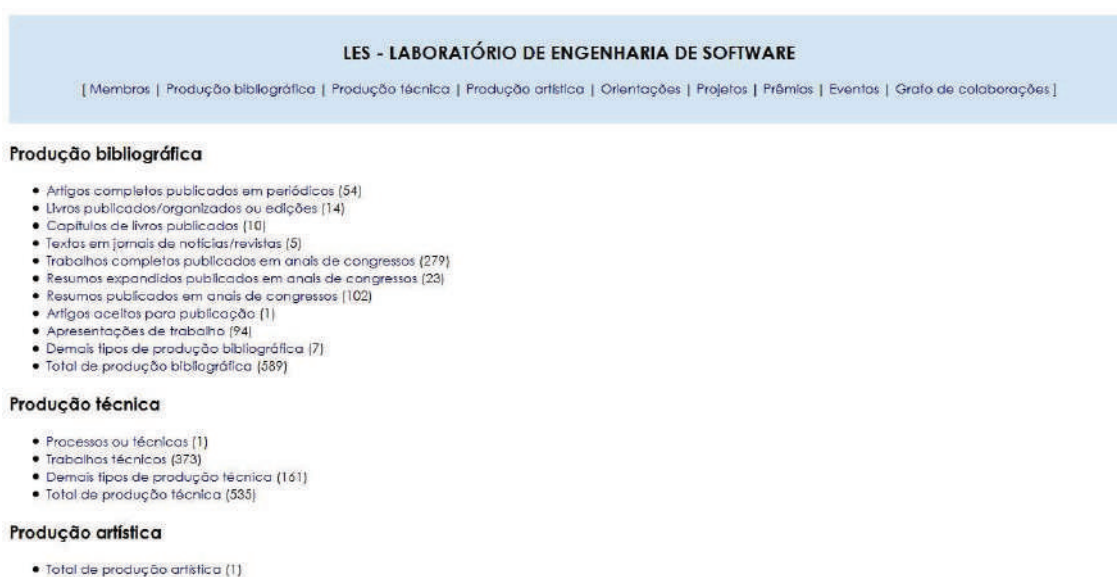


Figura 15. Quantidade de produções do LES identificadas com o ScriptLattes em 19/07/2016

A ferramenta identificou todas as produções do grupo no período de 2010 a 2015 que possuíam no mínimo: autor, título, local e ano de publicação e as classificaram corretamente conforme informações registradas no currículo lattes de cada pesquisador. Somadas as quantidades de produções, algumas foram apresentadas como não acessíveis, ou seja, não foram cadastradas corretamente pelos pesquisadores em seu currículo lattes, o que gerou dados de produções em categorias incorretas (ex. “apresentações de trabalhos” listadas na categoria “resumos publicados em anais”). Essas inconsistências foram deduzidas dos resultados finais das produções analisadas.

Devido à formatação das produções bibliográficas, utilizou-se para análise: (i) artigos completos publicados em periódicos; (ii) textos em jornais de notícias/revistas; (iii) trabalhos completos publicados em anais de congressos; (iv)

resumos expandidos publicados em anais; e (v) resumos publicados em anais de congressos.

Como apresentado na Figura 16, foram identificados:

- 54 artigos completos publicados em periódico e todos foram considerados para análise;
- 5 textos em jornais de notícias/revistas foram identificados. Destes 3 apresentaram-se como não acessíveis e não possuíam informações suficientes para identificação. Desta forma, somente 2 foram considerados para a análise;
- 279 trabalhos completos publicados em anais de congressos foram identificados. Destes 12 apresentaram-se como não acessíveis. Nesta classe não acessíveis 9 não possuíam dados suficientes para identificação das produções e 3 possuíam dados suficientes e foram identificados manualmente (2 do ano de 2010 e 1 do ano de 2011). Desta forma, 270 trabalhos completos publicados em anais de congressos foram considerados para a análise;
- 23 resumos expandidos publicados em anais de congressos foram identificados. Destes 2 estavam na classe não acessíveis. Nesta classe 1 não possuía informações suficientes para identificação e 1 do ano de 2012 possuía informação suficiente e foi identificado manualmente. Desta forma, 22 resumos expandidos publicados em anais de congressos foram considerados para a análise;
- 102 resumos publicados em anais de congressos foram identificados. Destes 85 estavam como não acessíveis. Deles 83 se referiam a participações em mesas redondas e não foram considerados na análise e somente 2 (1 do ano de 2012 e 1 do ano de 2015) possuíam informações suficientes e foram identificados manualmente. Foram considerados 19 resumos publicados em anais de congressos para a análise; e

- Ao final, 367 produções bibliográficas ou publicações foram consideradas para realização das análises do LES. Elas estão listadas conforme o Quadro 3 e disponível para consulta no link¹².

Observa-se com esse processo que os pesquisadores devem ter comprometimento ao cadastrar suas publicações na Plataforma Lattes o que permite uma maior confiabilidade dos dados e uma base sólida, o que contribui para a melhoria da qualidade das pesquisas que a utiliza como fonte de dados.

De posse das 367 publicações, os metadados do grupo foram extraídos e analisados. Como já relatado, todos os dados que serão apresentados são do período de 2010 a 2015.

¹² http://www.mestrado.web2405.uni5.net/Publicacoes_do_Grupo.xhtml

Artigos Completos							
Identificados e Analisadas	2010	2011	2012	2013	2014	2015	
54	= 4	+ 6	+ 7	+ 16	+ 10	+ 9	

Textos em Jornais / Revistas							
Identificados	2010	2011	2012	2013	2014	2015	Não acessível
5	= 0	+ 0	+ 0	+ 0	+ 0	+ 2	+ 3
Analisados	2010	2011	2012	2013	2014	2015	
2	= 0	+ 0	+ 0	+ 0	+ 0	+ 2	

Trabalhos completos publicados em anais de congressos							
Identificados	2010	2011	2012	2013	2014	2015	Não acessível
279	= 60	+ 43	+ 50	+ 38	+ 30	+ 46	+ 12
Analisados	2010	2011	2012	2013	2014	2015	
270	= 62	+ 44	+ 50	+ 38	+ 30	+ 46	

Resumos expandidos publicados em anais de congressos							
Identificados	2010	2011	2012	2013	2014	2015	Não acessível
23	= 2	+ 4	+ 1	+ 6	+ 4	+ 4	+ 2
Analisados	2010	2011	2012	2013	2014	2015	
22	= 2	+ 4	+ 2	+ 6	+ 4	+ 2	

Resumos publicados em anais de congressos							
Identificados	2010	2011	2012	2013	2014	2015	Não acessível
102	= 6	+ 1	+ 2	+ 5	+ 1	+ 2	+ 85
Analisados	2010	2011	2012	2013	2014	2015	
19	= 6	+ 1	+ 3	+ 5	+ 2	+ 2	

Figura 16. Quantidade e tipos de publicações analisadas

A Figura 17 apresenta a quantidade de publicações do LES, por tipo. De posse desses dados, os líderes e pesquisadores do LES podem verificar se precisam investir na produção de pesquisas de determinados tipos, com o objetivo de melhorar os índices. Observa-se que poucos textos foram publicados em jornais/revistas, o que pode caracterizar uma dificuldade do grupo de publicar nesses veículos. No ano de 2010 a 2015 foram publicados somente 2 textos em periódicos/revistas, que totalizam 0,5% das publicações. A maior quantidade de publicações do LES são trabalhos completos publicados em congressos 270-73,6%. Artigos completos são 54-14,7%. Resumos expandidos (22-6,0%) e resumos publicados em anais de congressos (19-5,2%) totalizam 41-11,2% das publicações.

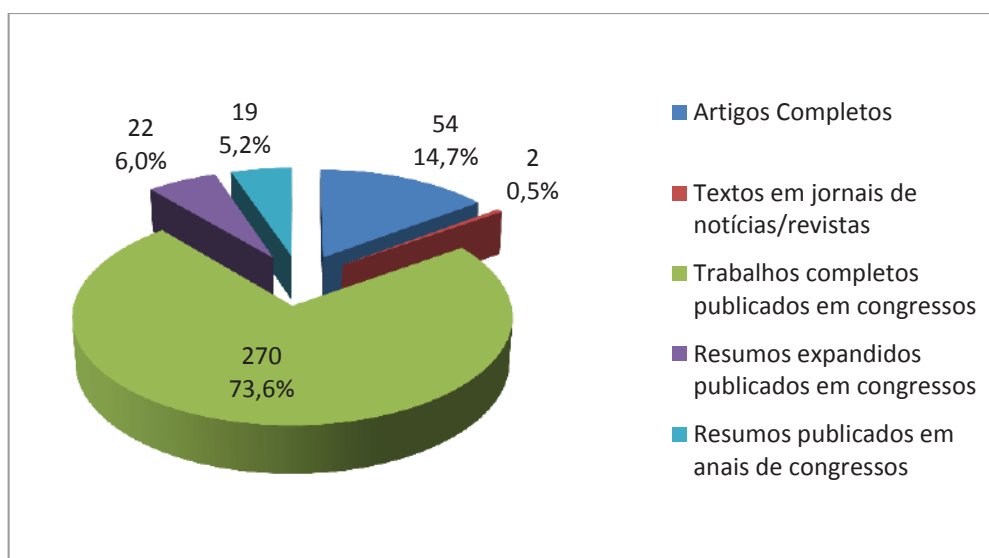


Figura 17. Quantidade de publicações do LES por tipo

A Figura 18 apresenta a quantidade de publicações por pesquisador do LES. Com esse gráfico pode-se identificar quais os pesquisadores que produzem pesquisas no laboratório. No caso do LES é possível observar que E. Almeida, do grupo RISE possui 115 publicações, o que representa 31,33% das publicações do LES, entre 2010 e 2015. Ele é o pesquisador que mais publicou entre 2010 a 2015 no laboratório, seguido por M. Mendonça (SoftVis) com 55 publicações, C. Chaves (Aside) com 42, I. Machado 41 e V. Santos 41, o que representa 48,77% das publicações.

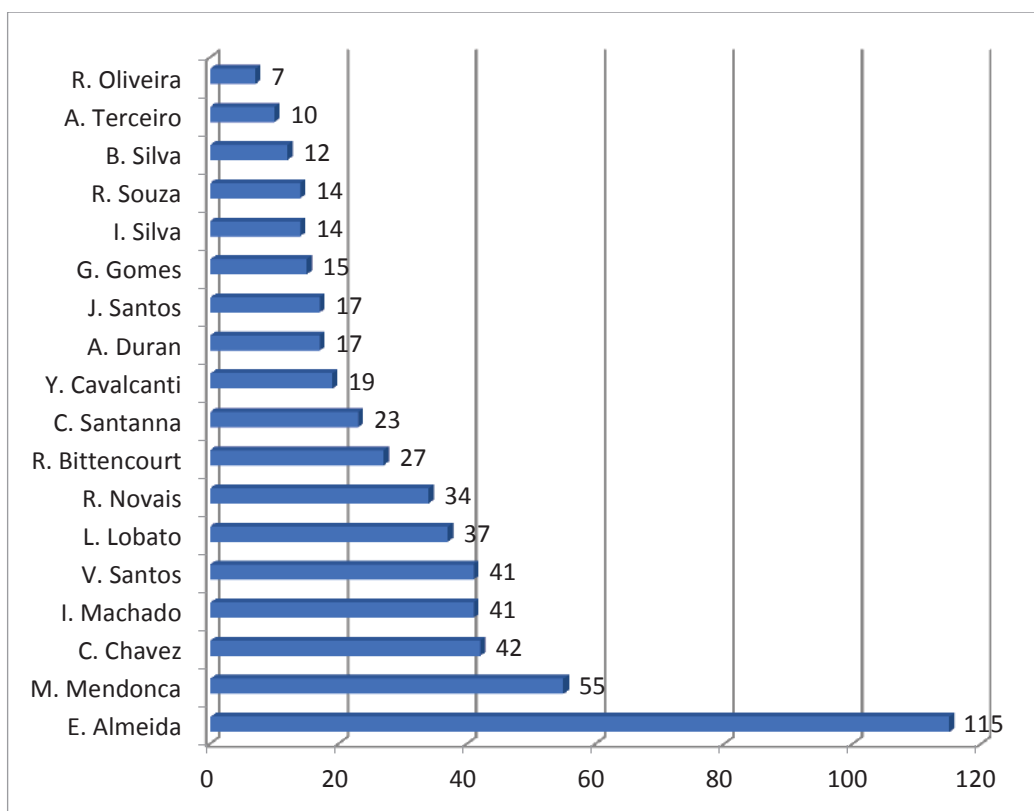


Figura 18. Quantidade de publicações por pesquisador do LES

A Figura 19 apresenta o grafo de colaboração dos pesquisadores, com peso. O peso se refere ao valor numérico associado a cada aresta. Elas mostram com quantos outros pesquisadores do grupo o pesquisador teve publicações e o peso está relacionado à quantidade de publicações.

Esse grafo foi extraído do relatório gerado pela ferramenta ScriptLattes. Ele serve para identificar como os pesquisadores se relacionam entre si. Importante para a criação e diversificação de novas pesquisas e projetos, bem como captação de recursos para a instituição e engajamento dos alunos em projetos, o que promove uma interação entre as áreas e disseminação do conhecimento. É possível observar por meio do grafo que todos os pesquisadores analisados colaboraram pelo menos 1 vez com outro pesquisador do laboratório. Manoel Mendonça colaborou com 11 pesquisadores e Eduardo Almeida com 10 pesquisadores, eles são os pesquisadores que mais se relacionam dentro do LES.

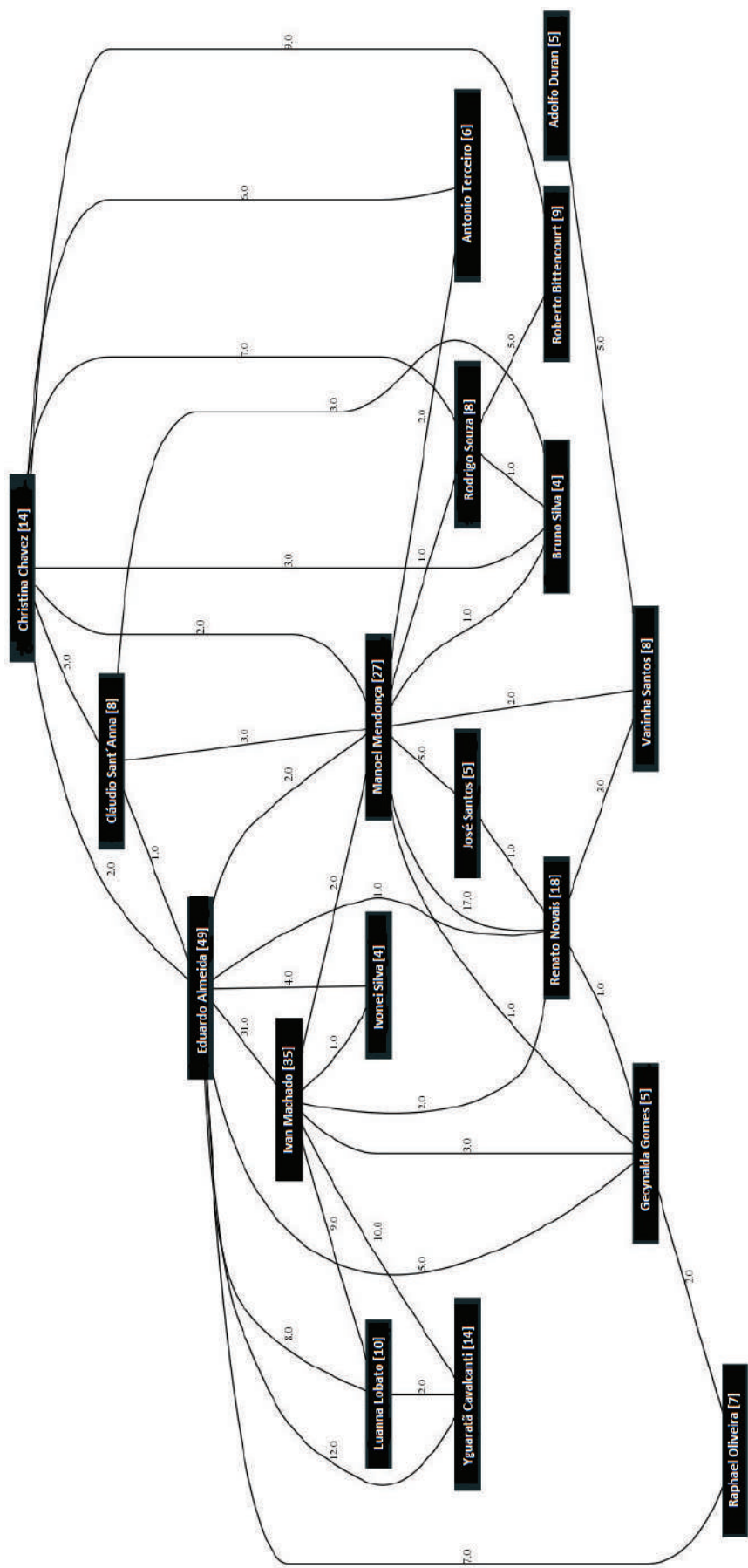


Figura 19. Grafo de colaboração dos pesquisadores do LES com peso

Finalizada a identificação do grupo e dos trabalhos desenvolvidos por seus pesquisadores, inicia-se a atividade de selecionar as publicações do grupo.

5.1.3 Atividade 3 – Selecionar as publicações do grupo

Para esta atividade foram reutilizados e adaptados do MS da área:

(I) Os critérios de inclusão:

- Publicações relacionadas com a área de MSR. Considerou-se relacionadas com a área de MSR todas as publicações que citaram ou realizaram mineração de artefatos de software para obter informações e/ou métricas;
- Publicações pertencentes ao domínio de ES; e
- Publicações desenvolvidas por pesquisadores do LES.

(II) Os critérios de exclusão:

- Publicações que não estão relacionados com a área de MSR no domínio de ES.
- Resumos de tutoriais, *surveys*, estudos experimentais secundários, *challenge* e *showcase*.

(III) As questões de pesquisa:

- Questão Geral (QG): Quais são as metas atuais abordadas pelas publicações relacionadas à MSR no LES?
- Questão Específica (QE)1. Quais são os principais propósitos, foco e objeto de análise, identificadas nas publicações?
- QE2. Quais são os tipos de fontes de dados suportadas por MSR e como elas estão evoluindo?
- QE3. Quais são as avaliações empíricas realizadas?

Para simplificar o entendimento, daqui por diante, para o texto “publicações relacionadas à área de MSR” será mencionado somente “publicações da área de MSR”.

Para a seleção das publicações do LES, consideradas relevantes para o estudo, foi utilizado o processo de triagem descrito na Figura 20. Este processo foi sugerido na descrição do MASG para facilitar o processo de rastreabilidade das publicações. A figura descreve brevemente o que foi realizado em cada atividade/filtro do processo. Além disso, apresenta a

quantidade de publicações selecionadas após a aplicação de cada atividade/filtro. As publicações selecionadas estão disponíveis no link¹³.

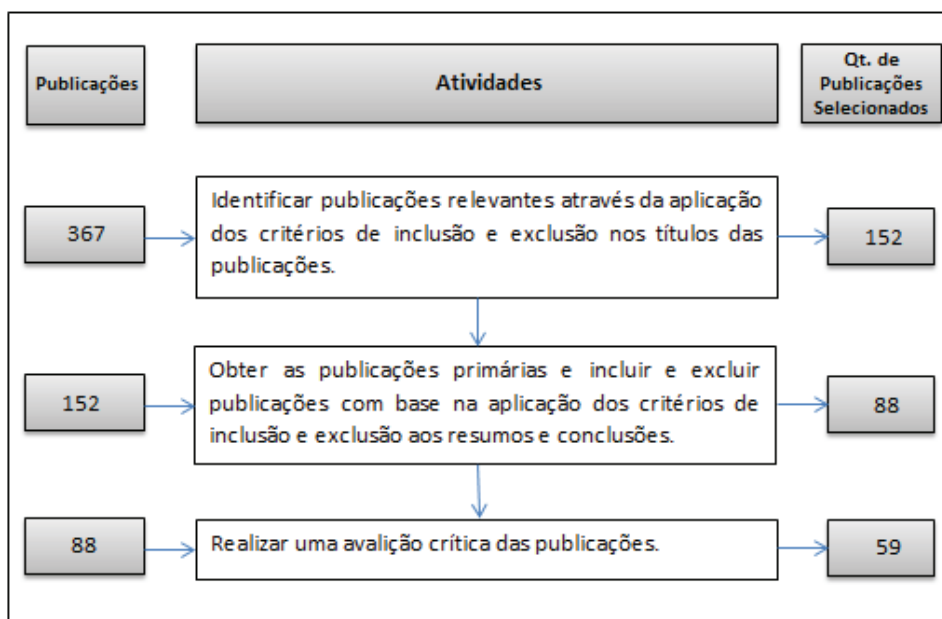


Figura 20. Triagem das publicações do LES

A Figura 21 apresenta a proporção das publicações do LES entre 2010 e 2015 na área de MSR e das demais áreas de conhecimento do laboratório. Observa-se que 59-16% das publicações são da área de MSR e 308-84% das demais áreas de conhecimento do laboratório. Esta quantidade de publicações na área de MSR é relevante. Vale salientar que o LES possui diversas linhas de pesquisa em ES e a área de MSR não aparece na especificação das linhas de pesquisa de nenhum dos grupos pertencentes a ele.

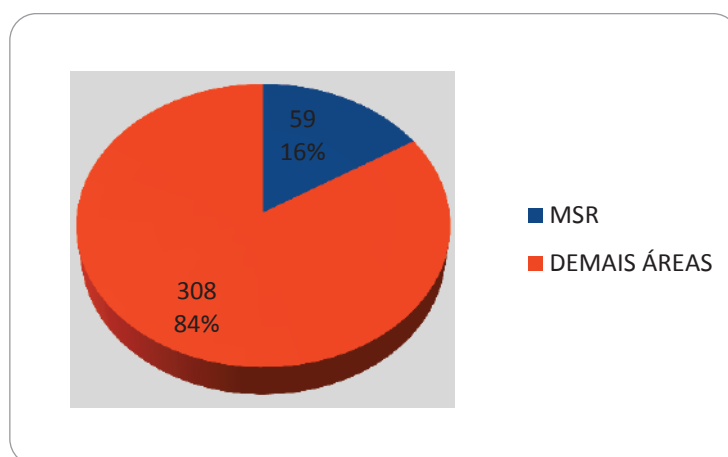


Figura 21. Proporção das publicações do LES entre 2010 e 2015 na área de MSR e das demais áreas de conhecimento

¹³ http://www.mestrado.web2405.uni5.net/Publicacoes_MSR_identificadas.xhtml

5.1.4 Atividade 4 – Extrair os dados das publicações do grupo

Esta **Atividade 4 – Extrair os dados das publicações do grupo** – teve como entrada a taxonomia, produzida no MS da área de MSR (Farias *et al.*, 2016) e atualizada neste estudo, utilizada para classificar e organizar os dados das publicações do LES. Essa taxonomia está disponível em endereço eletrônico¹⁴. O Quadro 7, disponível no Capítulo 4, foi utilizada como referência para extração dos dados de cada publicação selecionada.

Três pesquisadores foram envolvidos na extração dos dados, com o objetivo de melhorar a qualidade da extração, a revisão foi feita por pares. Dois pesquisadores leram cada produção e fizeram as classificações por meio de acordos. Caso os desentendimentos não fossem resolvidos por meio de discussões, foi solicitada a opinião de um terceiro pesquisador.

A classificação das publicações do grupo está disponível em endereço eletrônico¹⁵. Finalizada a classificação, a próxima atividade de análise dessa classificação das publicações do grupo foi realizada, como descrito na próxima subseção.

5.1.5 Atividade 5 – Analisar a classificação das publicações do grupo

Com base nos dados extraídos das 59 publicações selecionadas, os mapas do grupo ou resultados das análises dos dados puderam ser apresentados.

A Figura 22 apresenta a quantidade de publicações do LES na área de MSR. Dos 18 pesquisadores, observa-se que 12 realizam pesquisas nessa área. Nessa figura uma mesma publicação pode ter sido desenvolvida por mais de um pesquisador do LES.

¹⁴ http://www.mestrado.web2405.uni5.net/Taxonomia_Atualizada.xlsx

¹⁵ http://www.mestrado.web2405.uni5.net/Classificacao_Trabalhos.xlsx

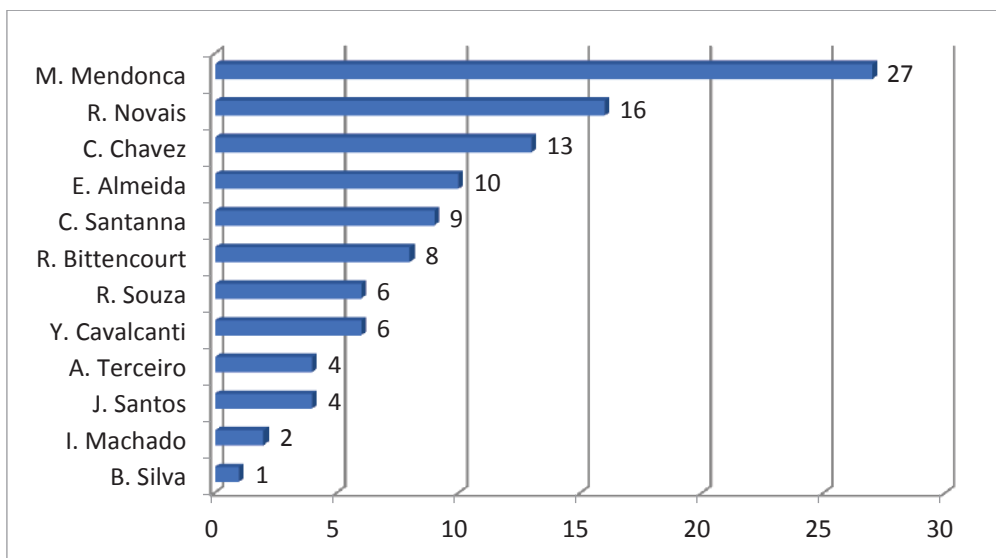


Figura 22. Publicações do LES por pesquisador na área de MSR

A Figura 23 apresenta a distribuição das publicações por grupo de pesquisa pertencente ao LES. Observa-se que dos cinco grupos desse laboratório três publicam nessa área.

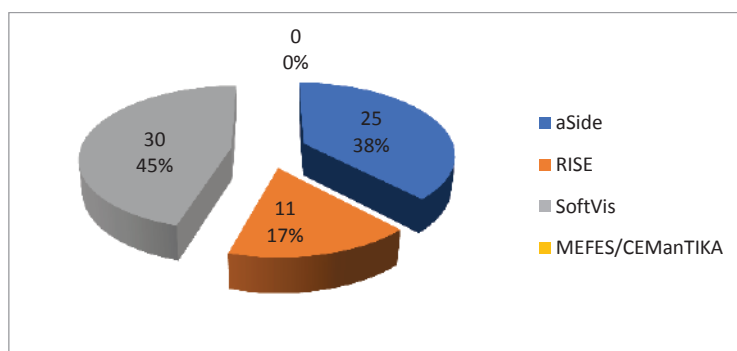


Figura 23. Publicações do LES por grupo de pesquisa na área de MSR

O grafo de colaboração com peso entre os grupos é apresentado na Figura 24. O peso, ou valor numérico, mostra a quantidade de publicações que um grupo colaborou com outro. Vale salientar que a mesma publicação pode ter sido produzida por pesquisadores dos 3 grupos. Percebe-se que há pouca colaboração entre os pesquisadores dos grupos que desenvolvem pesquisas em MSR no LES. Essa colaboração, conforme análise das publicações, foi realizada em 6 das 59 publicações do LES na área de MSR. Esse cenário mostra que há muitas possibilidades de interação entre os grupos do LES que podem ser exploradas.

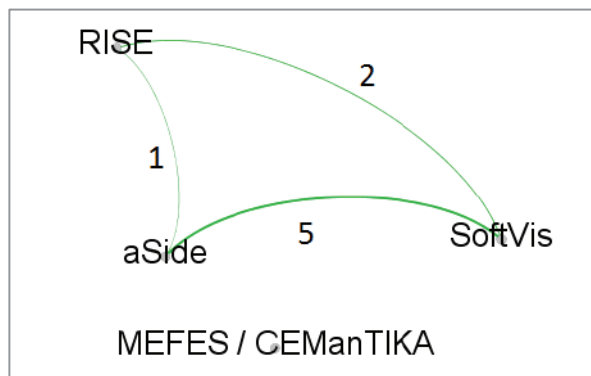


Figura 24. Grafo de colaboração com peso dos grupos do LES na área de MSR

A Figura 25 apresenta o grafo de colaboração dos pesquisadores do LES que tiveram publicações conjuntas na área de MSR. Esse grafo pode ajudar os pesquisadores a encontrarem outros parceiros que tenham objetivos comuns de pesquisa, com os quais ainda não tenham interagido. Observa-se que as fortes ligações são entre os pesquisadores do mesmo grupo (ex. R. Bittencourt, C. Chavez e R. Souza do aSide; M. Mendonça e R. Novais do SoftVis; e Y. Cavalcanti e E. Almeida do RISE).

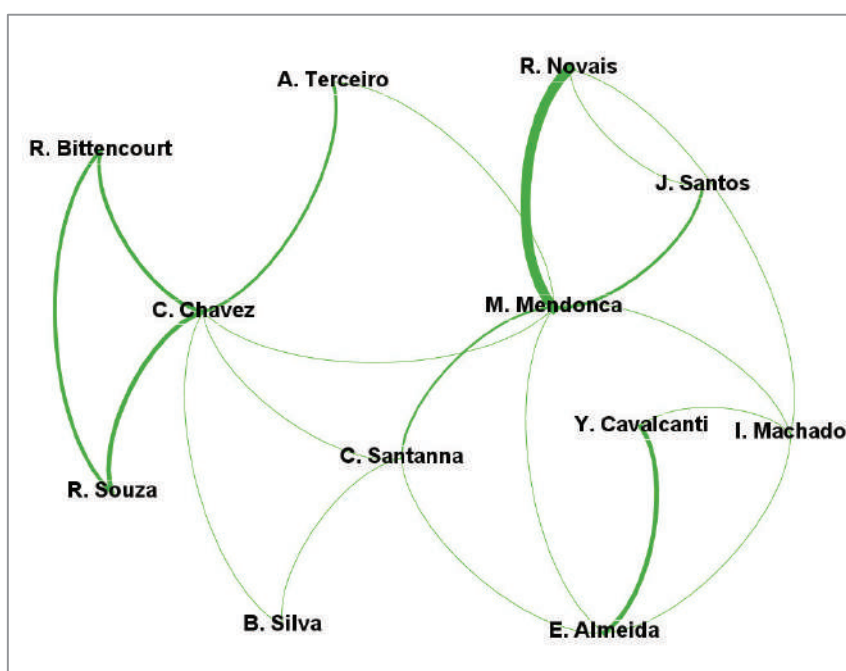


Figura 25. Grafo de colaboração dos pesquisadores do LES que tiveram publicações na área de MSR

A Figura 26 apresenta o grafo de colaboração de todos os pesquisadores que tiveram publicações conjuntas com os pesquisadores selecionados do LES na área de MSR.

Nele pode-se observar os pesquisadores internos e externos que contribuíram em pesquisas do LES na área de MSR. Esse grafo por cores ajudar a identificar os grupos (*clusters*) de pesquisadores com base em seu grau de interação.

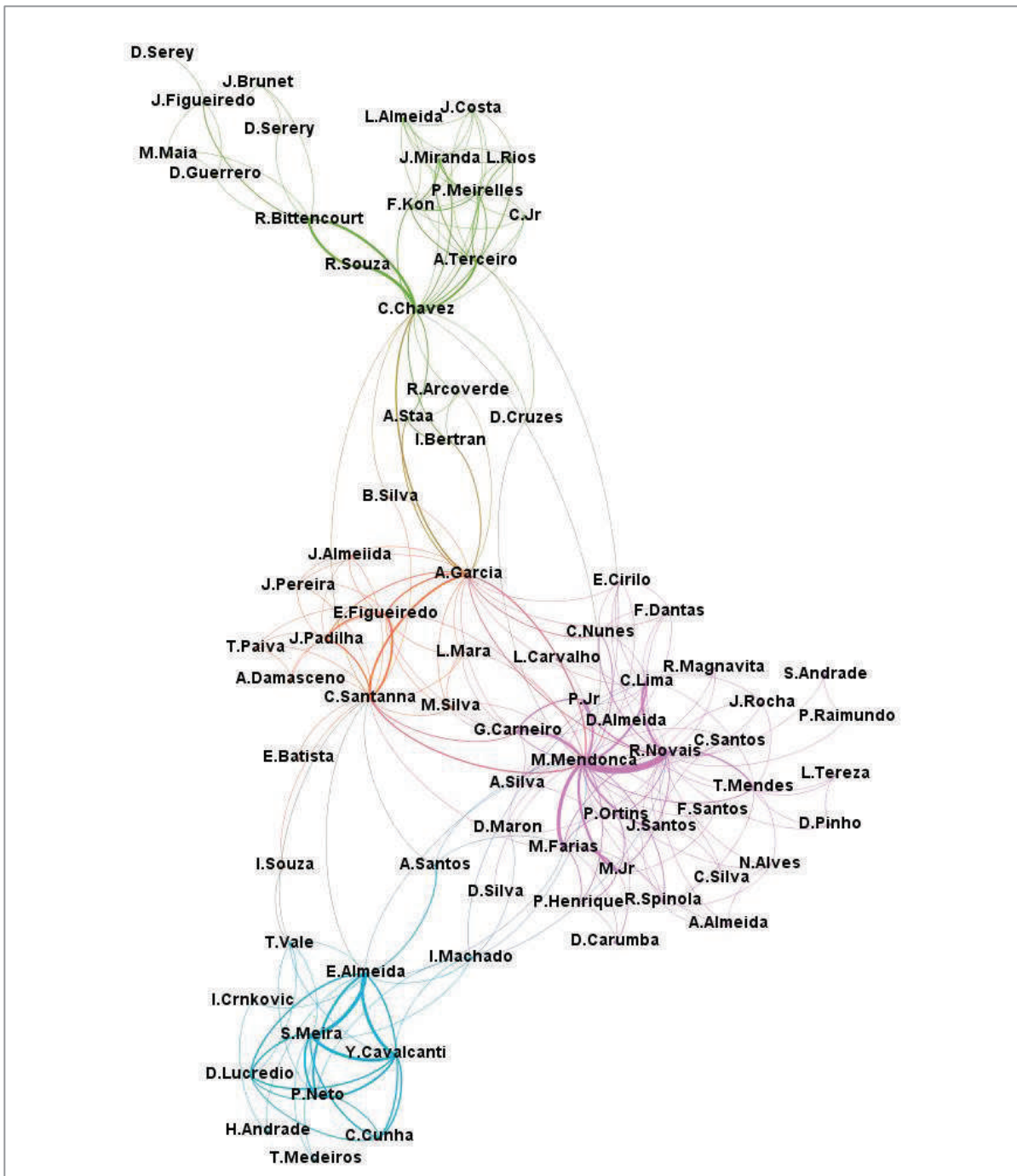


Figura 26. Grafo de colaboração dos pesquisadores que tiveram publicações conjuntas com os pesquisadores do LES na área de MSR

A Figura 27 apresenta a quantidade de publicações do LES na área de MSR por pesquisador (autores e coautores) que tiveram, no mínimo, 3 publicações. Nessa figura são apresentados os pesquisadores do LES e também pesquisadores que são autores ou coautores das publicações. A lista completa dos pesquisadores está disponível no Apêndice B.

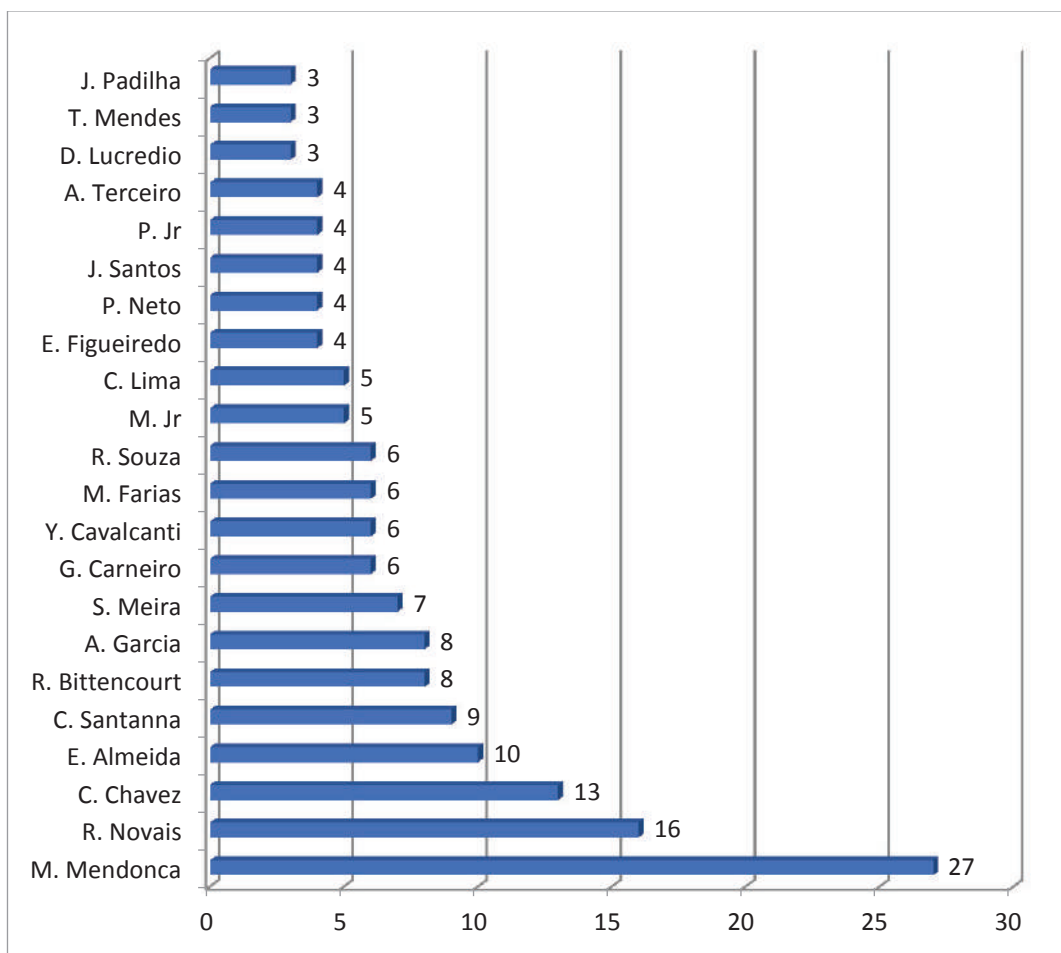


Figura 27. Quantidade de publicações do LES na área de MSR por pesquisador (mínimo 3) (autores/coautores)

A partir deste ponto, algumas figuras/mapas do LES estarão apresentadas em inglês para facilitar as comparações com as figuras/mapas que serão extraídas do MS de Farias *et al.*, 2016 e que serão apresentadas na próxima seção,

As respostas às questões de pesquisa seguem abaixo, fundamentadas pelas análises dos dados extraídos das publicações do LES. Vale salientar que nas análises feitas, uma publicação pode ter mais de um propósito, foco e objeto de análise. As classificações

dessas facetas seguem a taxonomia atualizada da área de MSR disponível em endereço eletrônico¹⁶.

QE1. Quais são os principais propósitos, foco e objeto de análise, identificados nas publicações?

Com relação ao propósito, pode-se dizer que de acordo com a Figura 28, 46-54,1% das publicações do LES utilizam o propósito de “compreensão” de algum tipo de artefato de software e dessas publicações, conforme dados das análises realizadas, 21-45,6% utilizam o recurso de visualização de software para apoiar esse propósito. Os propósitos “compreensão” e “identificação” são utilizados pela maioria das publicações analisadas, o que totalizam 81,2%. Poucas pesquisas foram realizadas, pelo LES, com o propósito de “melhoria” 1-1,2%, “classificação” 2-2,4% e “predição” 1-1,2%. Nenhuma publicação foi identificada com o propósito de “localização”. “Caracterização” foi o propósito utilizado por 6 publicações e “avaliação” também por 6, e totalizam 14,2% das publicações na área de MSR.

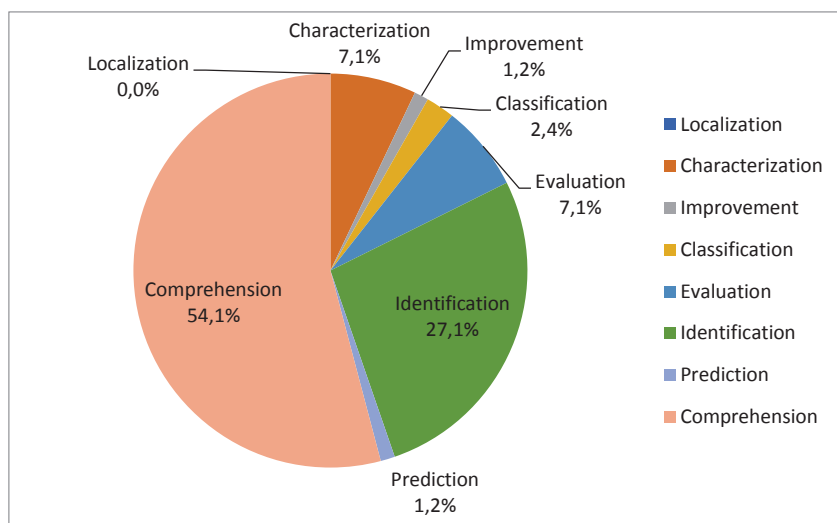


Figura 28. Propósitos utilizados pelas publicações do LES na área de MSR

Com relação ao foco de interesse das publicações, a Figura 29 apresenta essa faceta, juntamente com a percentagem de publicações em cada elemento dessa categoria. Nota-se que vários focos não são abordados pelo laboratório na área de MSR, como: “reuso de software”, “características do software”, “consumo de energia” e “esforço de desenvolvimento”. “Contribuição e comportamento de desenvolvedor” 14-19,4% é o foco

¹⁶ http://www.mestrado.web2405.uni5.net/Taxonomia_Atualizada.xlsx

mais explorado pelos pesquisadores do laboratório, seguido de “*smells*” 11-15,3%, “defeitos” 10-13,9% e “evolução de software” 10-13,9%, que totalizam mais da metade das publicações 45-62,5%. Os focos “métricas de software” 1-1,4%, “clones de código” 1-1,4%, “dívida técnica” 2-2,8%, “complexidade do software” 2-2,8% e “mudanças” 3-4,2% ainda são pouco explorados pelo LES. Os focos “qualidade dos artefatos de software” 5-6,9%, “Desenvolvimento de processos” 4-5,6%, “manutenabilidade” 5-6,9% e “características da arquitetura” 4-5,6%, totalizam 18-25% das publicações.

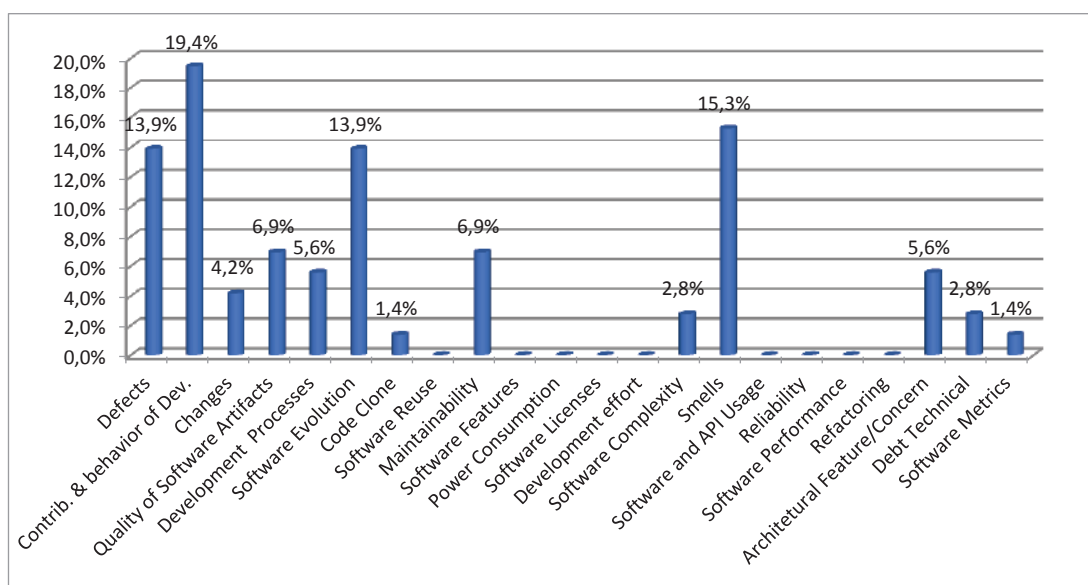


Figura 29. Focos utilizados pelas publicações do LES na área de MSR

Com relação ao objeto de análise, observa-se na Figura 30 que a maioria das publicações utilizam os objetos de análise “código” 31-39,7%, “dados de *commits*” 14-17,9% e “relatórios de bugs” 12-15,4%, presentes em 73% das publicações.

Com relação ao objeto de análise, pode-se dizer que, nenhuma publicação utiliza os seguintes objetos de análise: “texto em documentos”, “apps móveis”, “stack trace”, “dados de testes”, “third-party library” e “*microblogging*”. Durante as análises, apesar de identificadas várias publicações do LES na área de testes de software, nenhuma utilizou a Mineração de Repositório de Software para analisar o objeto de análise “dados de testes” ou qualquer outro objeto relacionado.

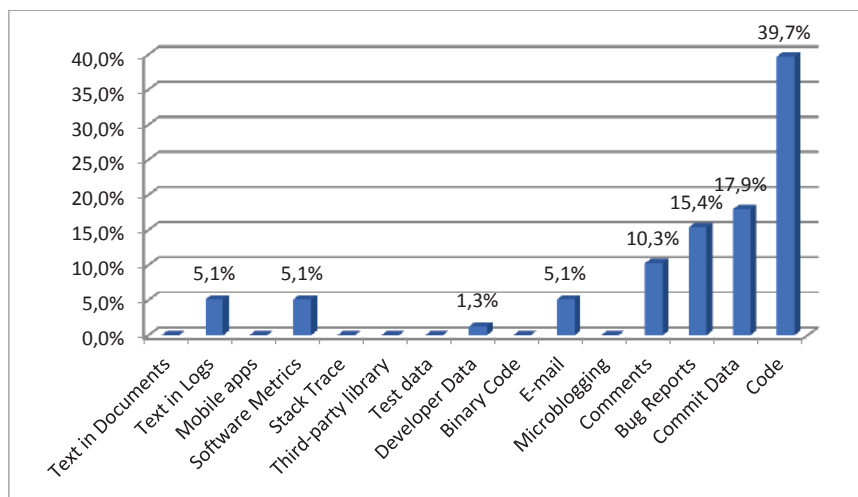


Figura 30. Objeto de análise utilizado pelas publicações do LES na área de MSR

Das publicações que utilizam o objeto de análise “e-mails” 4-5,1%, duas delas utilizam o objeto de análise “e-mails” e as outras duas utilizam “e-mails” combinado com “dados de *commits*”. Essas publicações são do mesmo pesquisador. Os objetos de análise “textos em *logs*” 4-5,1%, “métricas de software” 4-5,1% e “dados de desenvolvedores” 1-1,3% foram pouco explorados nas publicações.

As figuras apresentadas anteriormente mostram uma visão individual de cada faceta (propósito, foco e objeto de análise) e o gráfico de bolhas da Figura 32 apresenta uma visão combinada das facetas duas a duas. Observa-se que o propósito “compreensão”, o mais utilizado, foi explorado com todos os focos que foram utilizados por publicações do LES, exceto com o foco “clone de código”.

O propósito “identificação” foi o segundo mais utilizado nos estudos e explorado com todos os focos utilizados por publicações do LES, exceto com “mudanças” e “métricas de software”. A análise do objeto “código” foi feita com todos os focos utilizados por publicações do laboratório, exceto com “mudanças” e “desenvolvimento de processos”. O objeto de análise “dados de *commits*” também foi utilizado com todos os focos utilizados por publicações do LES, exceto com “métricas de software”, “clone de código”, “complexidade do software” e “*smells*”.

O objeto de análise “relatórios de bugs” também foi muito utilizado com o foco “defeitos”, bem como o objeto de análise “código” com o foco “evolução do software”.

Em resposta à questão, **QE2: “Quais são os tipos de fontes de dados suportadas por MSR e como elas estão evoluindo?”**, a Figura 31 apresenta a quantidade de publicações do LES na área de MSR por tipo de fonte de dados. Pode-se verificar que mais da metade das

publicações utilizam somente uma fonte de dados 50-84,8%, a saber, “código fonte” 30-50,8%, “BTS” 9-15,3%, “metadados de código fonte” 5-8,5%, “lista de e-mails” 2-3,4% e “logs” 4-6,8%.

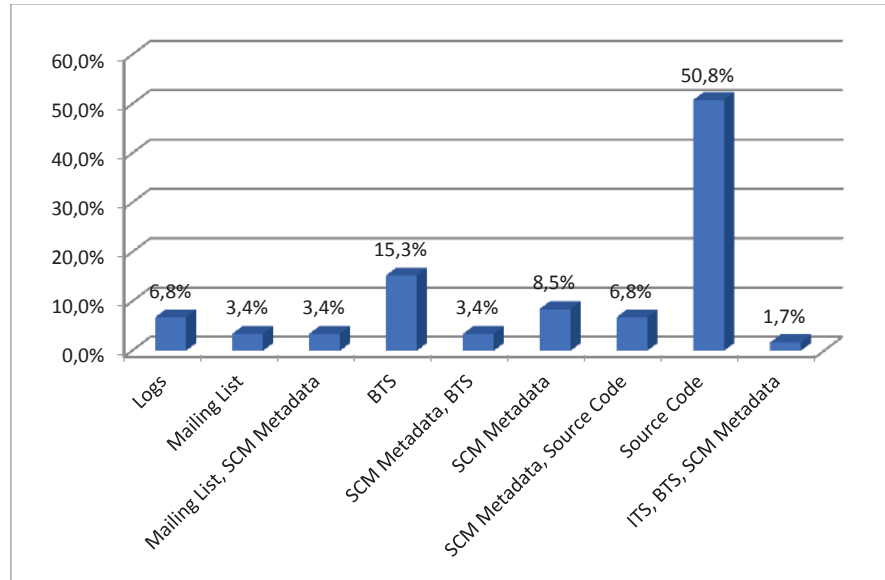


Figura 31. Quantidade de publicações do LES na área de MSR por tipo de fonte de dados

Observa-se também na Figura 31 que algumas fontes de dados foram utilizadas combinadas com outras, exemplo, “lista de e-mails” com “metadados de código fonte” e “metadados de código fonte” com “código fonte”, dentre outras. Os “metadados de código fonte” é a fonte de dados utilizada em todas as combinações. As publicações que utilizam mais de uma fonte de dados somam 9-15,3% do total. Das fontes de dados utilizadas pelas publicações da área de MSR do LES “código fonte” é a principal e foi utilizada em quase metade das publicações analisadas (34-57,6%).

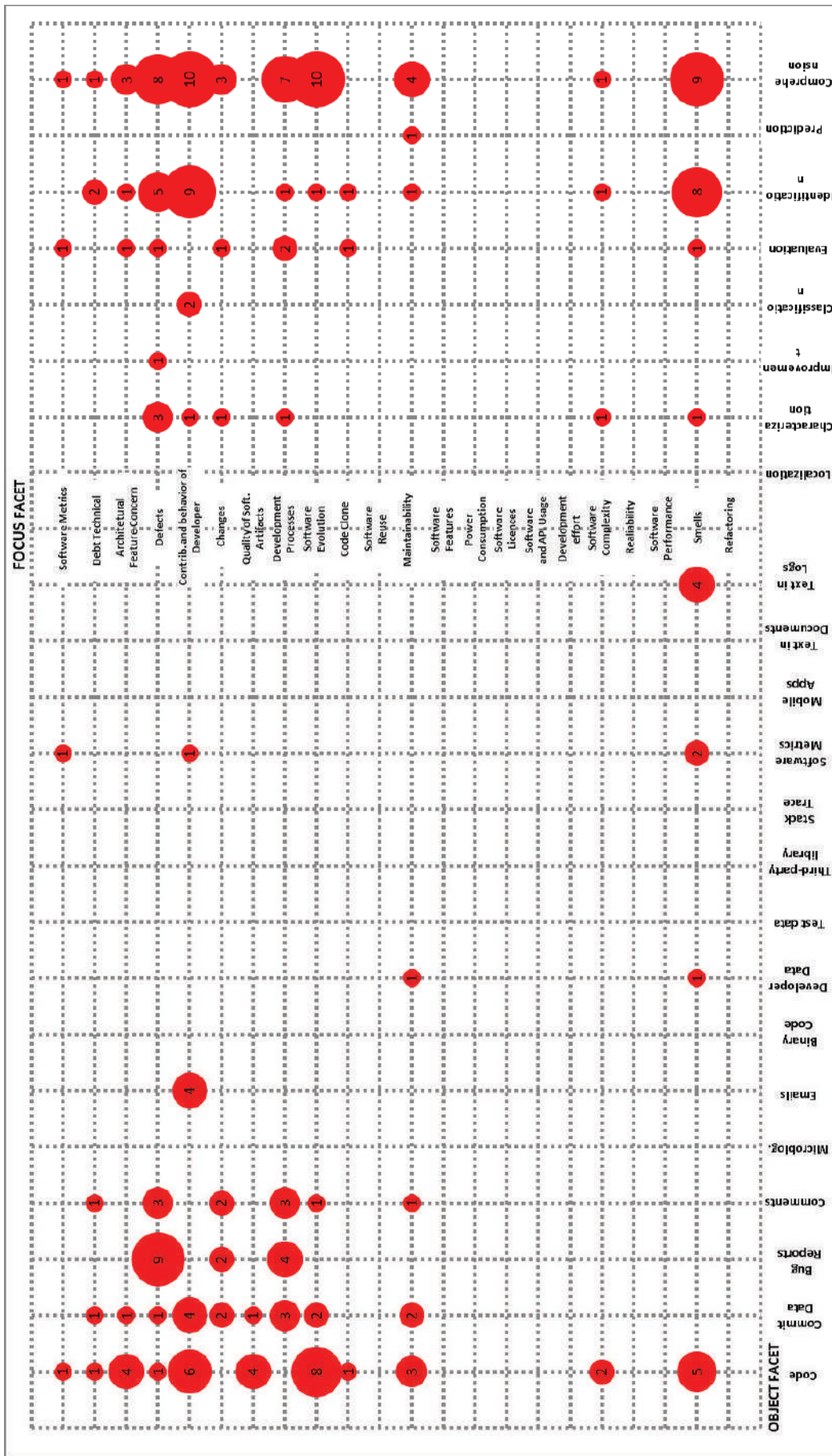


Figura 32. Gráfico de bolhas (Foco x propósito) da área de MSR do LES

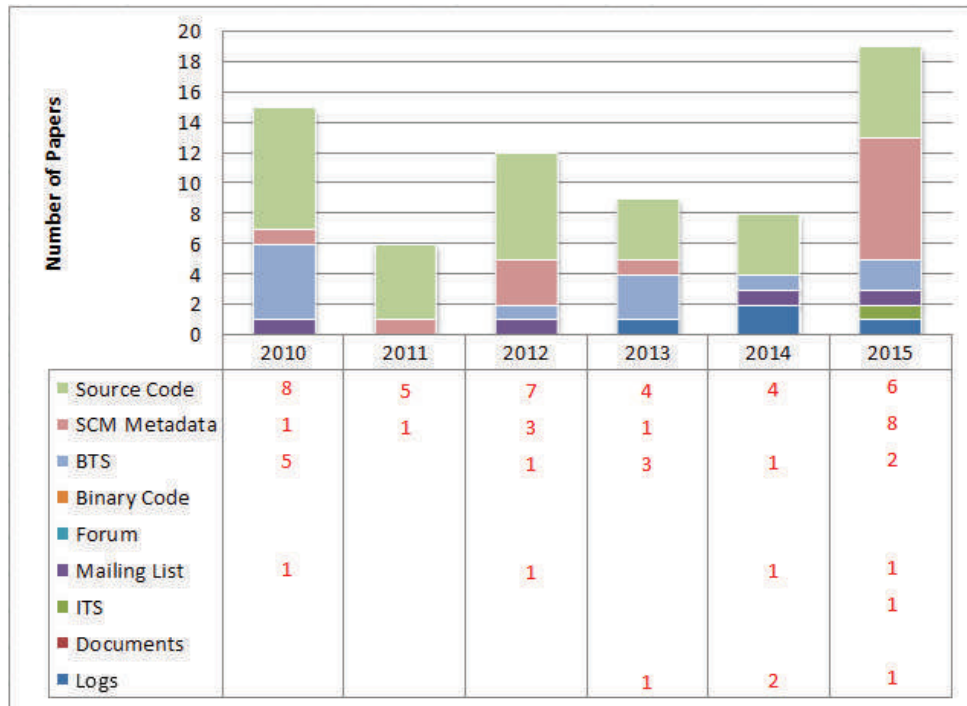


Figura 33. Fontes de dados utilizadas pelas publicações do LES na área de MSR por ano

Ainda em resposta à questão **QE2**, a Figura 33 apresenta a evolução das fontes de dados. Observa-se que há uma frequência maior, durante os anos, de publicações que utilizam as fontes de dados “código fonte”, “metadados de código fonte” e “BTS”. A fonte de dados “ITS” só foi utilizada no ano de 2015. A fonte de dados “lista de e-mails” também foi explorada e “logs” vem sendo explorada durante os anos desde 2013. Não foram identificadas publicações que utilizam as fontes de dados “fórum”, “documentos” e “código binário”.

A Figura 34 responde a **QE3: “Quais são as avaliações empíricas realizadas?”** Nessa figura observa-se que “estudo exploratório” 25-42,4% é o método de avaliação mais utilizado nas publicações do LES na área de MSR. Em 16-27,1% das publicações não foi realizado nenhum tipo de avaliação. Conforme as leituras dessas publicações do LES verificou-se que trata-se de trabalhos que apresentam ferramentas e abordam, somente, exemplos do seu uso ou descrição das suas funcionalidades. Nenhuma publicação utilizou “survey” para avaliação. As demais publicações utilizaram os seguintes tipos de avaliação: “estudo de viabilidade” 1-1,7%, “comparação com outras ferramentas” 3-5,1%, “experimento controlado” 7-11,9% e “estudo de caso” 7-11,9%, que totalizam 18-30,6% das publicações.

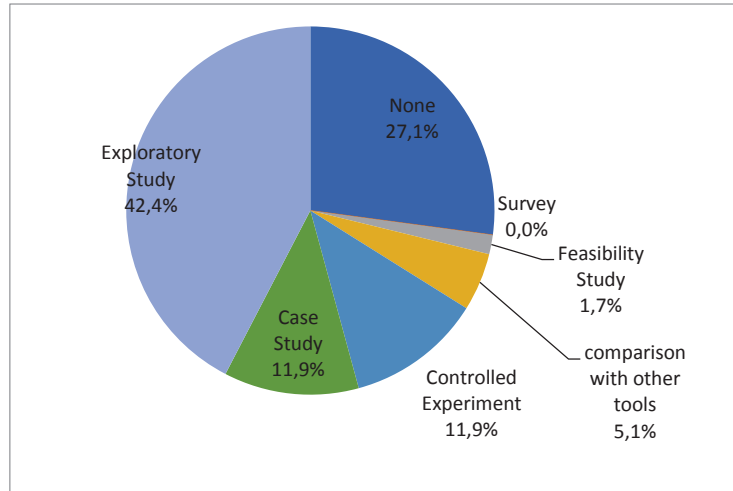


Figura 34. Métodos de avaliação utilizados pelas publicações do LES na área de MSR

A Figura 35 apresenta o gráfico de bolhas da combinação das facetas: propósito e método de avaliação. Essa figura evidencia a grande utilização de “estudo exploratório” com o propósito “compreensão” de algum objeto de análise (17 estudos). Geralmente as publicações que não utilizam métodos de avaliação utilizam o propósito “compreensão” para entender algum artefato de software (14 estudos), mas os propósitos “identificação” e “melhoria” também foram utilizados em publicações que não realizaram nenhum tipo de avaliação.

O propósito “compreensão” foi utilizado com todos os métodos de avaliação, exceto “*survey*” e “estudo de viabilidade”. O método de avaliação “*survey*” não foi utilizado por nenhuma publicação do LES na área de MSR.

A única publicação que utiliza o propósito de “melhoria” não fez nenhum tipo de avaliação e a única que utiliza o propósito de “predição” utiliza o método de avaliação “estudo de viabilidade”.

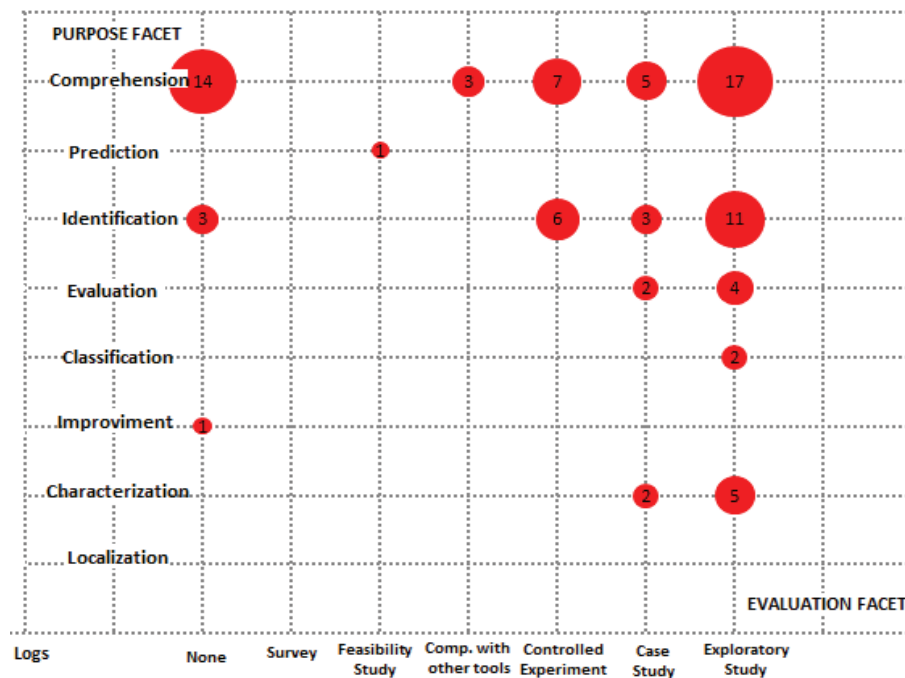


Figura 35. Gráfico de bolhas (Propósito x método de avaliação) do LES na área de MSR

A próxima subseção apresenta os resultados da atividade de análise e comparação dos mapas (da área e do laboratório).

5.1.6 Atividade 6 – Analisar e comparar os mapas

De posse dos mapas da área, obtidos do MS da área de MSR, e dos mapas do LES elaborados na atividade anterior, esta **Atividade 6 – Analisar e comparar os mapas** – apresenta a análise e comparação desses mapas.

Vale lembrar que foram analisadas 59 publicações do LES e 107 trabalhos da área de MSR e que essas publicações podem fazer referência a mais de um foco, propósito e/ou objeto de análise.

As comparações têm por objetivo identificar a posição do LES em relação à área de MSR apresentada no MS de Farias *et al.*, 2016.

A Figura 36 apresenta a comparação dos propósitos utilizados nas publicações do LES e dos trabalhos da área de MSR. Percebe-se que os propósitos das publicações do LES

acompanham os propósitos dos trabalhos da área de MSR. A maioria das publicações do laboratório (46-54,1%) utiliza o propósito “compreensão” de algum tipo de artefato de software, conforme as análises das classificações das publicações. Também percebe-se que as publicações com os propósitos “localização” e “melhoria” são pouco utilizadas pela área de MSR. O LES tem poucas publicações que utilizam o propósito de “predição”, o qual é o segundo mais utilizado pela área de MSR. O segundo propósito mais utilizado pelo LES é o de “identificação”. Os propósitos “classificação” e “avaliação” seguem a mesma proporção da área.

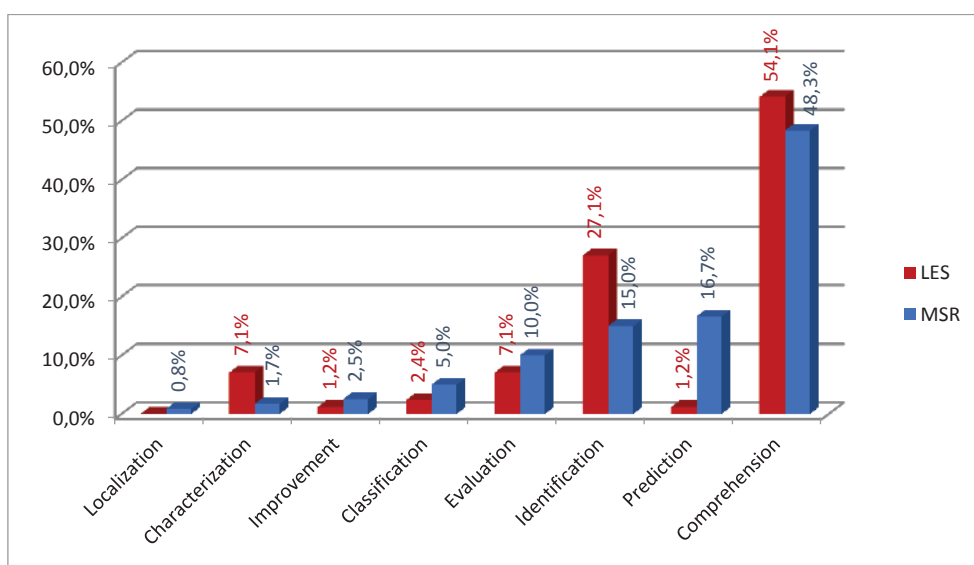


Figura 36. Comparação dos propósitos utilizados pelas publicações (MSR x LES)

Na Figura 37 observa-se que o foco “defeitos” é o mais explorado na área de MSR enquanto que o foco “contribuição e comportamento do desenvolvedor” é o mais explorado nas publicações do LES o que representa 14-19,4% das publicações desse laboratório. Conforme a análise das publicações, todas que estão relacionadas ao foco “contribuição e comportamento do desenvolvedor” avaliam o comportamento dos desenvolvedores dado um cenário, com relação a assuntos como: complexidade estrutural, sistema de representação preferido, dentre outros.

Os focos “características da arquitetura” 4-5,6%, “dívida técnica” 2-2,8% e “métricas de software” 1-1,4% não foram abordados nos trabalhos da área de MSR. Os focos “defeitos” 10-13,9%, “mudanças” 3-4,2% e “evolução de software” 10-13,9% são

classificados como os tópicos mais populares da área de MSR (Farias *et al.*, 2016) e também são explorados em publicações do laboratório.

Apesar de terem sido pouco abordados pela área de MSR, os focos “reuso de software”, “características de software”, “consumo de energia”, “licenças de software”, “esforço de desenvolvimento”, “*software* e *API usage*”, “confiabilidade”, “performance de *software*” e “refatoração”, não foram abordados nas publicações do LES, o que pode sinalizar lacunas para novas pesquisas do laboratório.

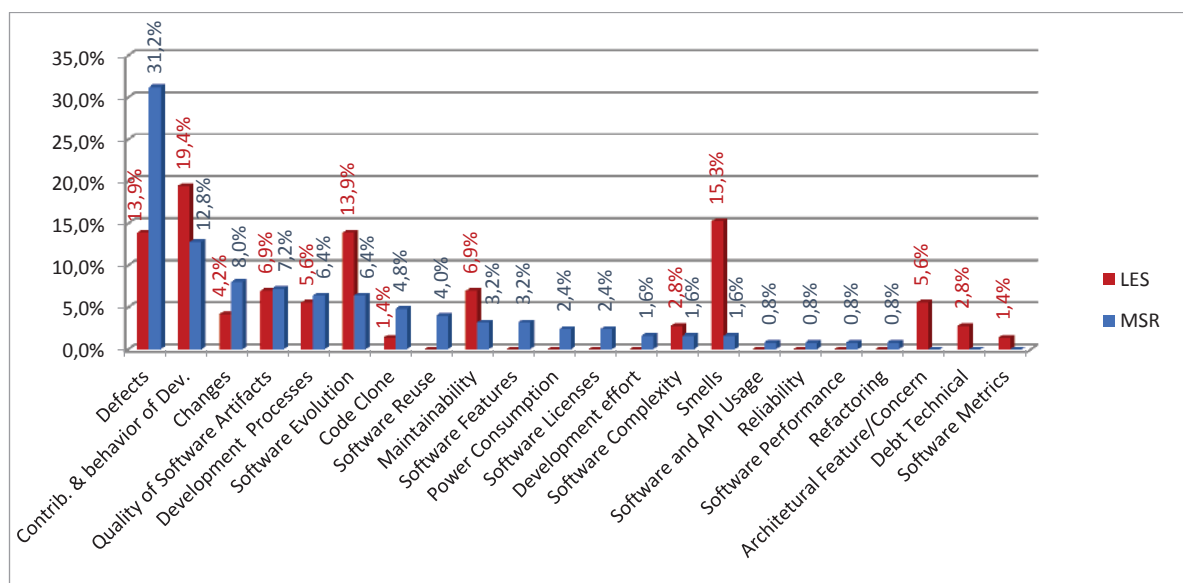


Figura 37. Comparação dos focos utilizados pelas publicações (MSR x LES)

Quanto ao objeto de análise, a comparação realizada na Figura 38 apresenta que os objetos de análise "código" (31-39,7%), "dados de *commits*" (14-17,9%), "relatórios de bugs" (12-15,4%) e "comentários" (8-10,3%) são os 4 objetos mais significativos e representam 65-83,3% de todos os estudos do LES, assim como, na área de MSR.

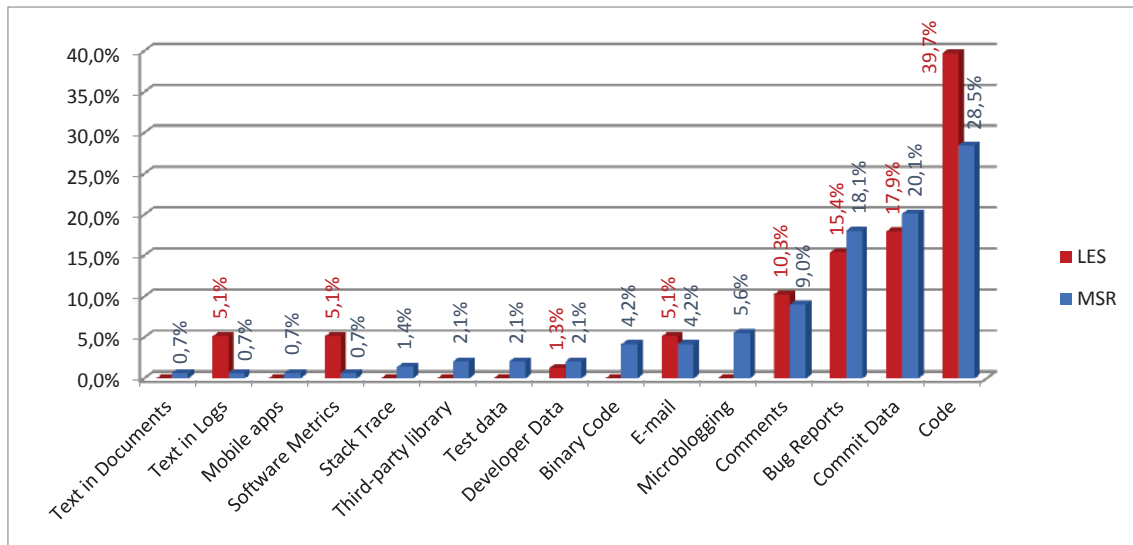


Figura 38. Comparação dos objetos de análise utilizados pelas publicações (MSR x LES)

A Figura 39 apresenta a sobreposição dos resultados dos gráficos de bolhas que combinam as facetas, propósito e método de avaliação da área de MSR e do LES. As bolas vermelhas representam o resultado das publicações do LES na área de MSR e as que têm somente o contorno preto representam o resultado dos trabalhos da área de MSR. Percebe-se que o propósito "caracterização" e o método de avaliação "estudo exploratório" foram utilizados em publicações do LES e não foi utilizado em nenhum trabalho da área de MSR, o que também acontece com as publicações do LES que utilizam o propósito "identificação" e não realizam nenhum tipo de avaliação. O propósito "melhoria" também foi utilizado somente pelo laboratório em publicações que não realizaram nenhum tipo de avaliação, bem como o propósito "compreensão" que utiliza o método de avaliação "comparação com outras ferramentas".

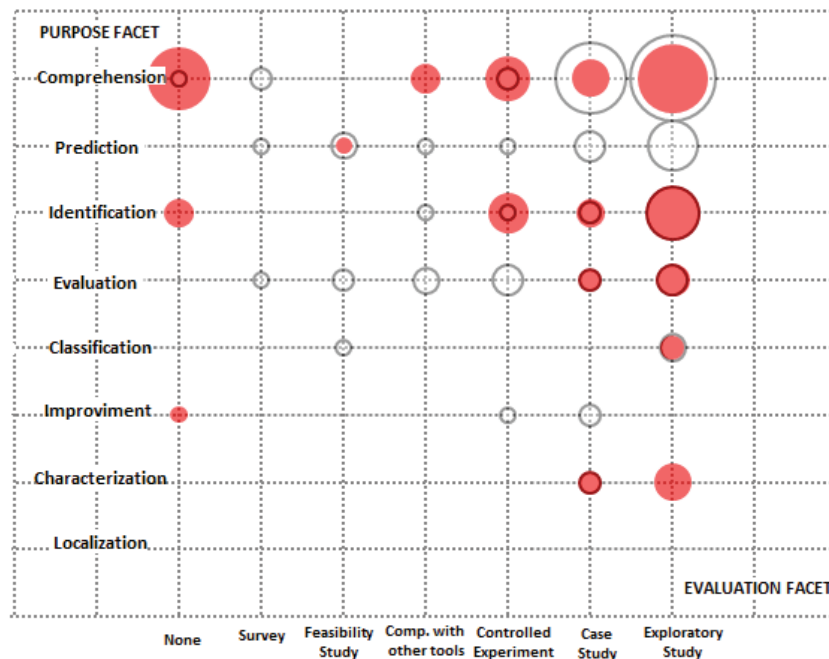


Figura 39. Sobreposição dos resultados dos gráficos de bolhas (propósito x método de avaliação) (MSR x LES)

O propósito "predição" só foi utilizado pelo LES em uma publicação. Diferentemente da área de MSR, que ele foi utilizado com todos os métodos de avaliação. O propósito "avaliação" também foi utilizado pela área de MSR com todos os métodos de avaliação e somente com "estudo de caso" e "estudo exploratório" pelo LES.

As publicações que usaram o propósito "compreensão" juntamente com o método de avaliação "estudo exploratório" são as mais frequentes, tanto na área como no laboratório.

Observa-se que os propósitos de "predição" e "avaliação" podem ser mais explorados pelo LES em publicações da área de MSR, visto que, são diversamente explorados pela área.

A Figura 40 apresenta a comparação das fontes de dados utilizadas nos estudos do LES e da área de MSR. Nesta figura uma mesma publicação pode pertencer a mais de uma categoria. Percebe-se que as publicações do LES estão alinhadas com as da área, no que diz respeito às fontes de dados mais utilizadas, que são: "código fonte", "metadados de código" e "BTS". As demais fontes, exceto "documentos", "ITS" e "logs" não são utilizadas em publicações do LES e são utilizadas pela área.

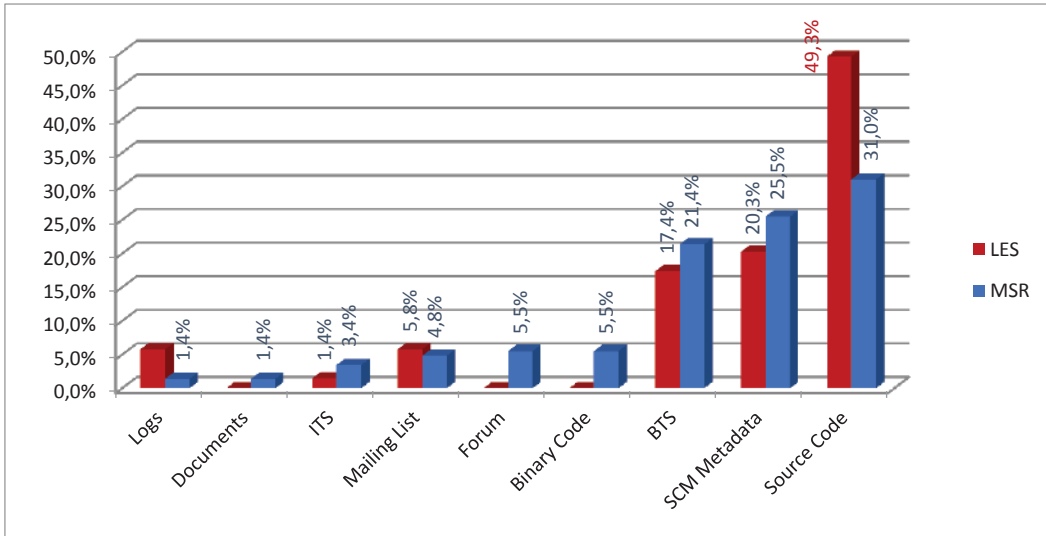


Figura 40. Comparação das fontes de dados utilizadas pelas publicações (MSR x LES)

Para melhor caracterizar as fontes de dados a Figura 41 apresenta a comparação das publicações que utilizaram mais de uma fonte de dados. Percebe-se que várias combinações de fontes de dados ainda não foram exploradas pelo LES, o que pode ser testado para desenvolver novas pesquisas, como: “BTS” e “código fonte”; “metadados de código fonte”, “código fonte” e “BTS”; e “metadados de código fonte”, “BTS” e “lista de e-mails”. Percebe-se que combinações de fontes de dados são uma das características da área de MSR. Essas combinações ajudam os pesquisadores a melhorar o entendimento do foco pretendido.

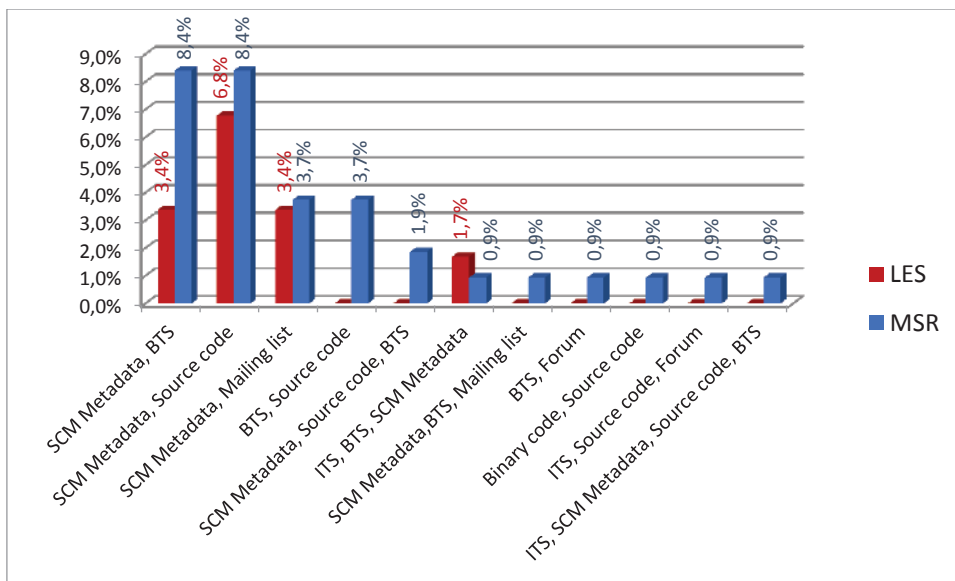


Figura 41. Comparação das publicações que utilizaram mais de uma fonte de dados (MSR x LES)

A Figura 42 apresenta na parte superior um gráfico que representa a quantidade de publicações do LES conforme a legenda de cores associada às fontes de dados relacionadas na parte inferior esquerda dessa figura. Na parte inferior direita têm-se, em valores numéricos, as quantidades de publicações da área (em preto) e do LES (em vermelho), para cada ano, entre 2010 e 2015.

Percebe-se que assim como na área de MSR, o LES tem explorado ao longo desses anos a fonte de dados “código fonte”, “metadados de código fonte” e “BTS”.

A fonte de dados “lista de e-mails” é utilizada em publicações tanto no laboratório como em trabalhos da área de MSR, enquanto “código binário”, “fórum” e “documentos” vêm sendo exploradas, apenas pela área de MSR. Desta forma, entende-se que o investimento em publicações que utilizem essas fontes de dados possa trazer a possibilidade de novas descobertas para o laboratório.

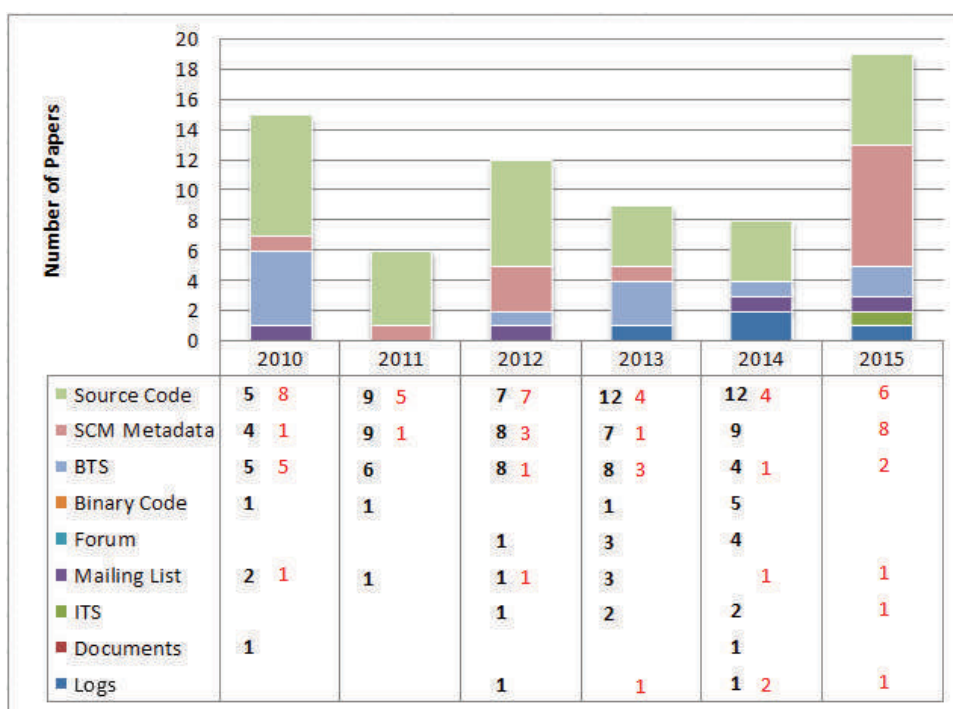


Figura 42. Quantidade de publicações por tipo de fontes de dados e ano (MSR x LES)

Na Figura 43 apresenta-se a comparação do LES com a área, em relação aos métodos de avaliação utilizados nas publicações. Entende-se que “estudo exploratório” e “estudo de caso” são os métodos de avaliação mais utilizados em trabalhos da área de MSR, já as publicações do LES utilizam com mais frequência os métodos de avaliação “estudo

exploratório”, “experimento controlado” e “estudo de caso”, bem como publicações que não utilizam nenhum método de avaliação.

Os métodos de avaliação são utilizados para melhorar a confiança e qualidade das pesquisas realizadas. As publicações analisadas que não possuem avaliação geralmente são publicações que apresentam ferramentas ou algumas de suas funcionalidades. O LES desenvolveu uma quantidade considerável de publicações que não possuem avaliação.

O método de avaliação “experimento controlado” é utilizado por publicações do laboratório em sua maioria com o foco “*smells*”.

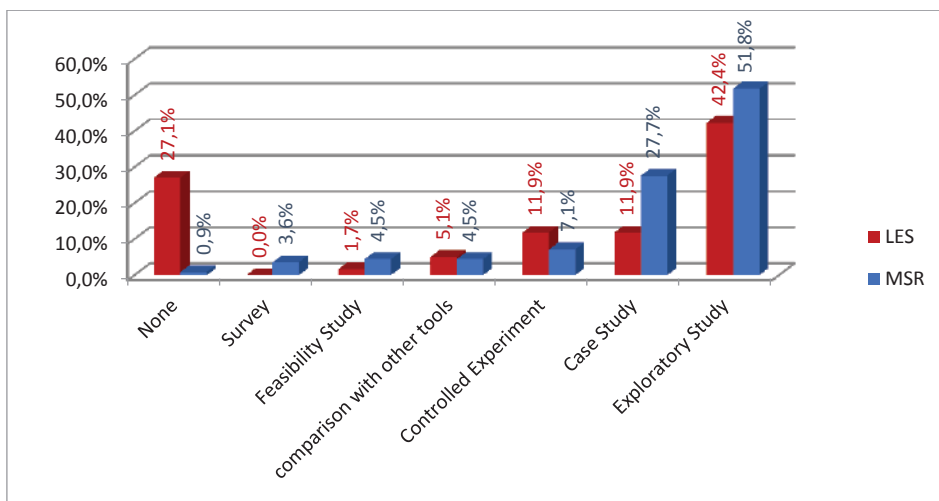


Figura 43. Comparação dos métodos de avaliação utilizados nas publicações (MSR x LES)

A Figura 44 apresenta uma visão geral dos mapas, com a sobreposição dos gráficos de bolhas que combinam: propósito e foco; e foco e objeto de análise. Nessa figura percebe-se que as publicações do LES se concentram nos objetos de análise “código” e “dados de *commits*”, bem como nos propósitos “compreensão” e “identificação”. Percebe-se com clareza que o foco “*smells*” é bem utilizado por publicações do LES e pouco por trabalhos da área de MSR.

Com a apresentação desta figura fica fácil identificar os objetos de análise inexplorados pelo LES e explorados pela área de MSR (ex. “*microblogging*”, “dados de teste”, “confiabilidade”, “reuso de *software*”, “consumo de energia”), bem como os focos abordados por publicações do LES e não explorados pela área de MSR (ex. “métricas de software”, “dívida técnica” e “características da arquitetura”). Essa figura permite que os

pesquisadores identifiquem os rumos que estão tomando as publicações da área e do laboratório, o que possibilita a tomada de decisão pelos líderes e pesquisadores do grupo.

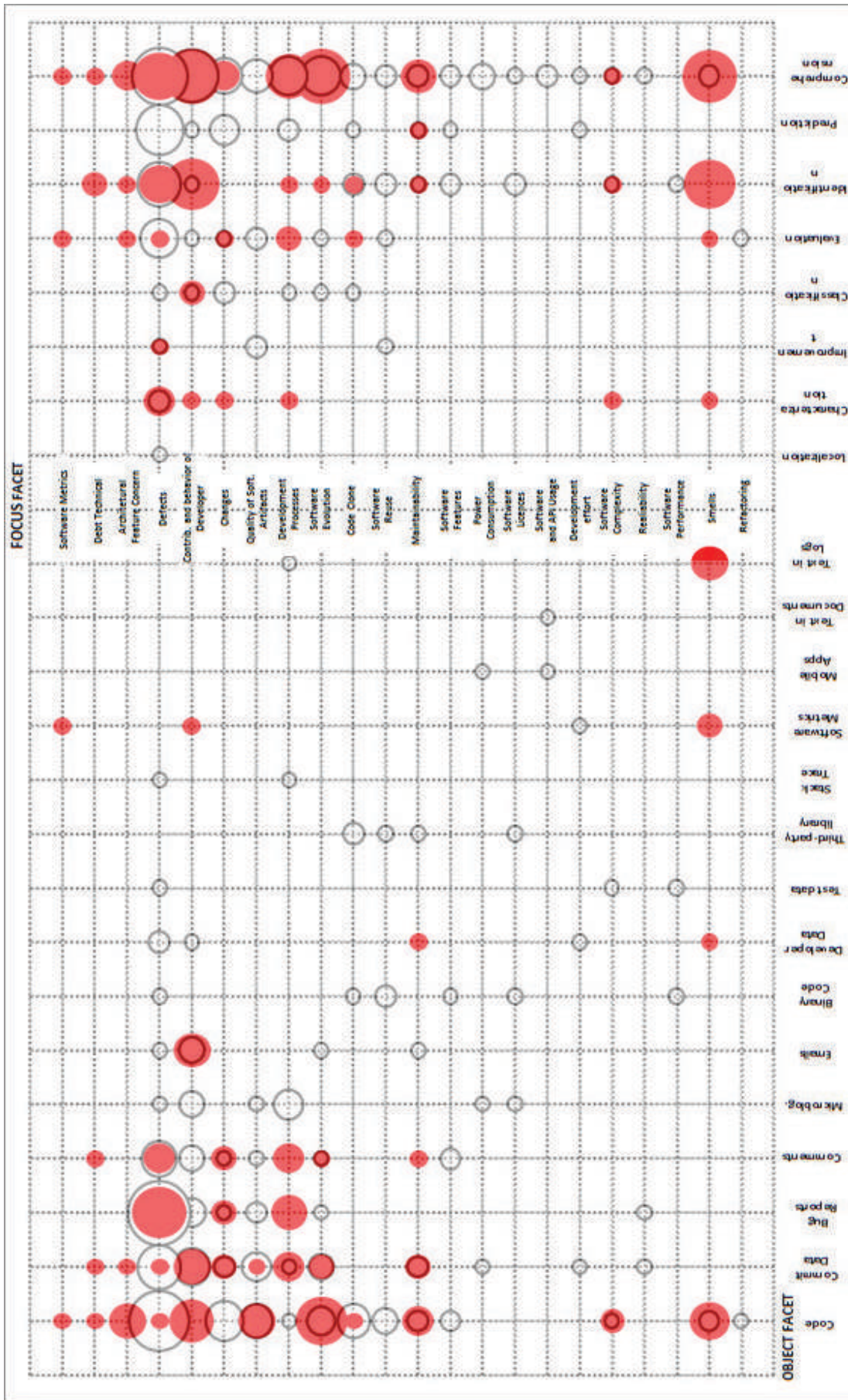


Figura 44. Sobreposição de resultados dos gráficos de bolhas (propósito x foco; foco x objeto de análise) (MSR x LES)

5.1.7 Atividade 7 – Entrevistar os pesquisadores do grupo

Nessa **Atividade 7 – Entrevistar os pesquisadores do grupo** – foram realizadas entrevistas individuais semiestruturadas com o objetivo de apresentar aos pesquisadores os resultados obtidos durante as atividades já executadas e capturar informações adicionais sobre sua visão em relação aos resultados, ao laboratório, e ao método, que não foram percebidas durante o processo e nas análises dos mapas. Esta atividade também visa identificar estratégias de pesquisa utilizadas pelos pesquisadores, como por exemplo, estratégias para captar recursos financeiros, melhorar a infraestrutura de pesquisa, melhorar recursos humanos, dentre outras.

Foi realizada uma entrevista piloto para verificar se o modelo proposto seria eficiente para obter os resultados desejados de apresentar com clareza e sem distorções os mapas construídos na Atividade 5 e 6. Também foram criados mapas de comparação construídos com os dados coletados de cada pesquisador para apresentação nas entrevistas, com o objetivo de colher as informações necessárias para concluir o estudo. Essas comparações apresentaram aos pesquisadores uma visão individualizada da posição das suas publicações na área de MSR em relação as do grupo e da área.

Durante essa entrevista foi percebido nos comentários e sugestões do entrevistado na etapa piloto, que a reorganização dos mapas apresentados no material poderia facilitar o entendimento dos entrevistados. Assim, os mapas dos resultados foram revistos, reestruturados, as cores de cada categoria criada foram padronizadas e as etapas da entrevista redefinidas. Toda essa reformulação foi realizada com o objetivo de dar um melhor entendimento aos entrevistados ao analisar o material utilizado para a entrevista.

As entrevistas foram realizadas com 8 pesquisadores doutores, pertencentes ao LES, que tiveram publicações na área de MSR identificadas no período de 2010-2015. Vale salientar que do total de 12 pesquisadores identificados para as entrevistas, somente 8 foram entrevistados devido à indisponibilidade de 4 deles.

As entrevistas tiveram duração de 19 minutos à 1 hora e 44 minutos, a depender da análise do material da entrevista feita pelo pesquisador. Alguns pesquisadores tiveram mais interesse em avaliar os resultados apresentados de forma minuciosa e fizeram contribuições com sugestões, críticas, elogios e comentários, enquanto outros se limitaram em analisar o material e responder aos questionamentos. Percebeu-se que o interesse do pesquisador pela área de interesse analisada (MSR), impactou neste quesito.

Todas as entrevistas foram gravadas e transcritas manualmente após a sua realização. Para transcrever cada entrevista utilizou-se, em média, 6 vezes o tempo da duração da gravação. Um modelo completo das entrevistas realizadas com os pesquisadores do LES encontra-se disponível no link¹⁷.

As entrevistas possibilitaram ter um melhor entendimento do grupo e colher dados importantes dos pesquisadores. Desta forma, o envolvimento dos pesquisadores no processo foi de fundamental importância para elaboração de uma atividade mais eficiente e uma coleta de dados mais confiável o que contribuiu com informações de valor e aplicáveis ao LES.

5.1.7.1 Dados quantitativos das entrevistas

A realização das entrevistas possibilitou a coleta de dados quantitativos que permitiram conclusões importantes sobre o LES. Na primeira etapa do questionário foram feitas perguntas sobre os mapas relacionados ao LES, ao pesquisador e à área de MSR. Também foi realizada uma pergunta relacionada à área de pesquisa do pesquisador.

A Figura 45 apresenta a quantidade de pesquisadores entrevistados que tem a área de MSR como área principal ou secundária de pesquisa. Resultado obtido com a **Pergunta 1**: “*Esta área de conhecimento (MSR) é a sua principal área de pesquisa ou somente área secundária? Explique.*”. Percebe-se que eles têm, em sua maioria, a área de MSR como área secundária (6-75%) de pesquisa. E utilizam essa área com o objetivo de encontrar evidências para responder perguntas de pesquisa para realizar publicações que se utilizam de grandes

¹⁷ http://mestrado.web2405.uni5.net/Modelo_Entrevista.pdf

quantidades de dado (em sua maioria presente em repositórios *open source*) e para analisar objetos de pesquisa de sua área principal. Observa-se que somente 2-25% dos pesquisadores têm a área de MSR como área principal de pesquisa. Áreas como visualização de software, reuso de software e linhas de produtos de software foram citadas pelos pesquisadores como áreas principais de pesquisa.

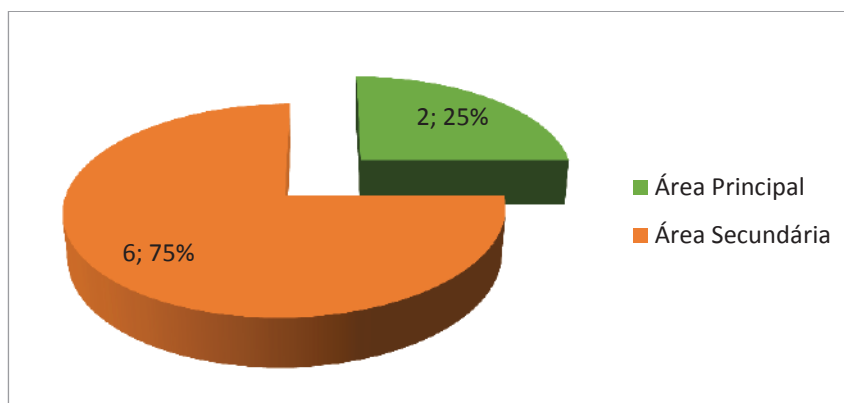


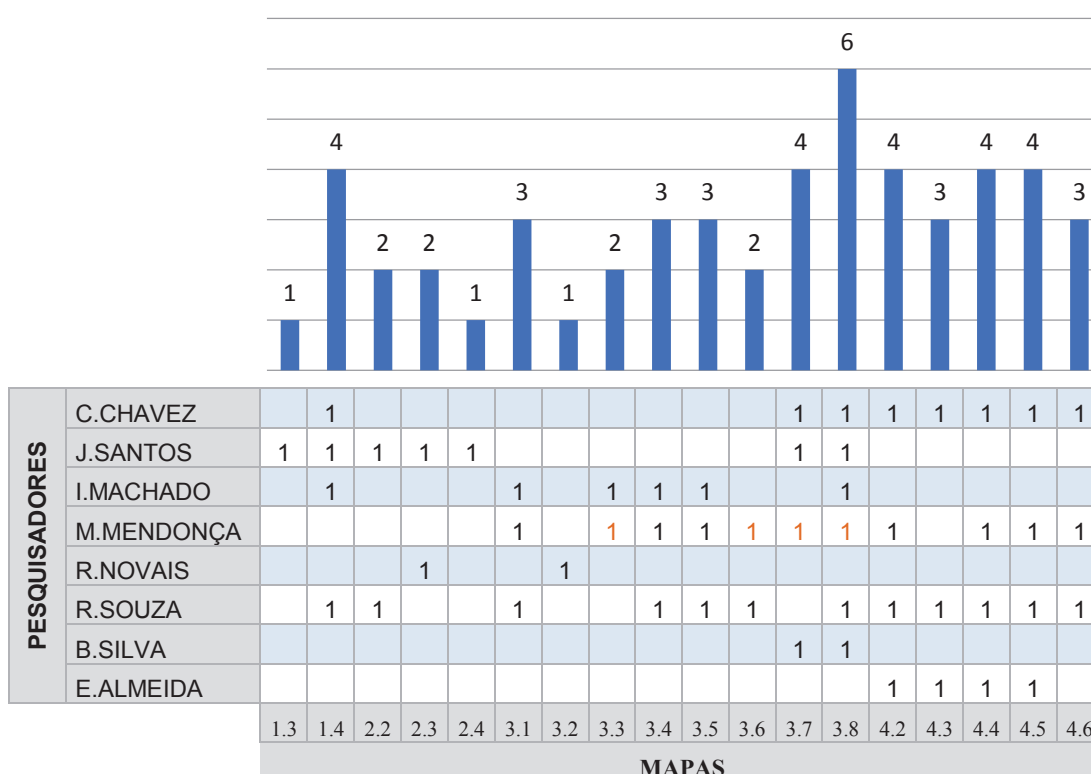
Figura 45. Quantidade de pesquisadores entrevistados que tem a área de MSR como área principal ou secundária de pesquisa

Com o resultado obtido com a **Pergunta 2**: “*Qual o seu principal objetivo e motivação para elaboração de estudos nesta área? Explique.*”. Observou-se que os pesquisadores realizam pesquisas na área de MSR com o objetivo de analisar grandes escalas de dados, em sua maioria, disponíveis em repositórios de softwares, bem como para analisar dados de vários projetos, geralmente, *open source* ou dados produzidos durante o desenvolvimento de estudos. Essa motivação se dá pelo interesse do pesquisador de encontrar evidência para responder hipóteses ou questões de pesquisa, muitas vezes, da sua área de conhecimento principal.

Com o resultado obtido com a **Pergunta 3**: “*As comparações dos mapas lhe permitem entender melhor a área de conhecimento e ter novas visões e direcionamentos para os seus estudos? Explique.*”. Percebeu-se que todos os pesquisadores (8-100%) responderam que as comparações ajudam. Desta forma, pode-se dizer que a **Atividade 6 – Analisar e comparar os mapas** é de grande importância para que o pesquisador tenha uma visão, de uma área de interesse, de um grupo e dos seus pesquisadores, bem como possibilita o

direcionamento de novas pesquisas a serem realizadas pelo grupo. Essas comparações de forma sistemática possibilitam também uma posterior comparação entre grupos e pesquisadores de outras instituições que venham aplicar o MASG.

A Figura 46 apresenta os mapas mais citados como úteis durante as entrevistas. Resultado obtido com a **Pergunta 4**: “Quais gráficos foram mais úteis e/ou interessante para você? Explique.”. Levou-se em consideração que os mapas citados como pouco úteis não deixam de ser úteis, desta forma, esses mapas também foram considerados na figura e identificados com a cor laranja.



- LEGENDA DOS MAPAS**
- 1.3. Quantidade de publicações por pesquisador do LES
 - 1.4. Grafo de Colaboração dos pesquisadores do LES com peso
 - 2.2. Publicações relacionadas à área de MSR por pesquisador do LES
 - 2.3. Grafo de colaboração dos pesquisadores do LES que tiveram publicações em MSR
 - 2.4. Quantidade de publicações MSR por pesquisador (mínimo 3) (autores/coautores)
 - 3.1. Fontes de Dados (MSR x LES)
 - 3.2. Sobreposição dos resultados das fontes de dados por ano (MSR x LES) (Evolução)
 - 3.3. Propósitos (MSR x LES)
 - 3.4. Objetos de Análise (MSR x LES)
 - 3.5. Focos (MSR x LES)
 - 3.6. Métodos de avaliação (MSR x LES)
 - 3.7. Sobreposição dos resultados dos gráficos de bolhas das facetar: propósito x método de avaliação (MSR x LES)

<p>3.8. Sobreposição de resultados dos gráficos de bolhas das facetas: propósito x foco x objeto de análise (MSR x LES)</p> <p>4.1. Lista das publicações relacionadas à área de MSR identificadas e analisadas</p> <p>4.2. Comparação de Fontes de Dados com os resultados do Pesquisador</p> <p>4.3. Comparação de Propósitos com os resultados do Pesquisador</p> <p>4.4. Comparação de Objetos de Análise com os resultados do Pesquisador</p> <p>4.5. Comparação de Focos com os resultados do Pesquisador</p> <p>4.6. Comparação de Métodos de Avaliação com os resultados do Pesquisador</p> <p>OBS: A numeração dos mapas e nomenclaturas corresponde às mesmas utilizadas no modelo de entrevista aplicado aos pesquisadores do LES (disponível no endereço eletrônico¹⁸)</p>
--

Figura 46. Mapas mais citados nas entrevistas

Conforme as análises das entrevistas, somente o grafo 2.3 (Grafo de colaboração dos pesquisadores MSR (autores/coautores)), foi citado por 4 pesquisadores como difícil de ser analisado pela sua forma de apresentação. Vale salientar que para esta dissertação foi feita uma nova apresentação desse grafo (Figura 26) com o objetivo de facilitar a análise e entendimento dos pesquisadores.

Na segunda etapa dos questionamentos das entrevistas foram colhidas informações sobre objetivos e estratégias de pesquisa dos pesquisadores. Os objetivos de pesquisa listados no questionário das entrevistas foram identificados por meio de discursões com especialistas.

A Figura 47 apresenta os resultados quantitativos da **Pergunta** realizada na segunda etapa do questionário: “*Marque com ‘X’ os objetivos que você utiliza e relate as estratégias que você utiliza para alcançá-los*”. Percebe-se que a maioria dos pesquisadores respondeu que tem o objetivo de pesquisa de “captação de recursos financeiros” (7-87,5%), “ampliação do grupo” (6-75%), “melhoria dos recursos humanos” (7-87,5%), “desenvolvimento de parcerias com outros grupos” (5-62,5%) e “identificação de novas oportunidades de pesquisa” (6-75%).

Dos pesquisadores envolvidos na entrevista 6-75% não têm o objetivo de “nucleação de novos grupos”. Os pesquisadores que responderam ter esse objetivo, o descreveu como uma consequência para o LES, pois alunos fazem doutorado e/ou mestrado e participam de grupo do LES e vão para outras instituições. Esses alunos, normalmente, criam

¹⁸ http://mestrado.web2405.uni5.net/Modelo_Entrevista.pdf

outros grupos que acabam a compartilhar interesses em comum com o seu grupo de origem que se interessa em nuclear o novo grupo formado para que ele se desenvolva.

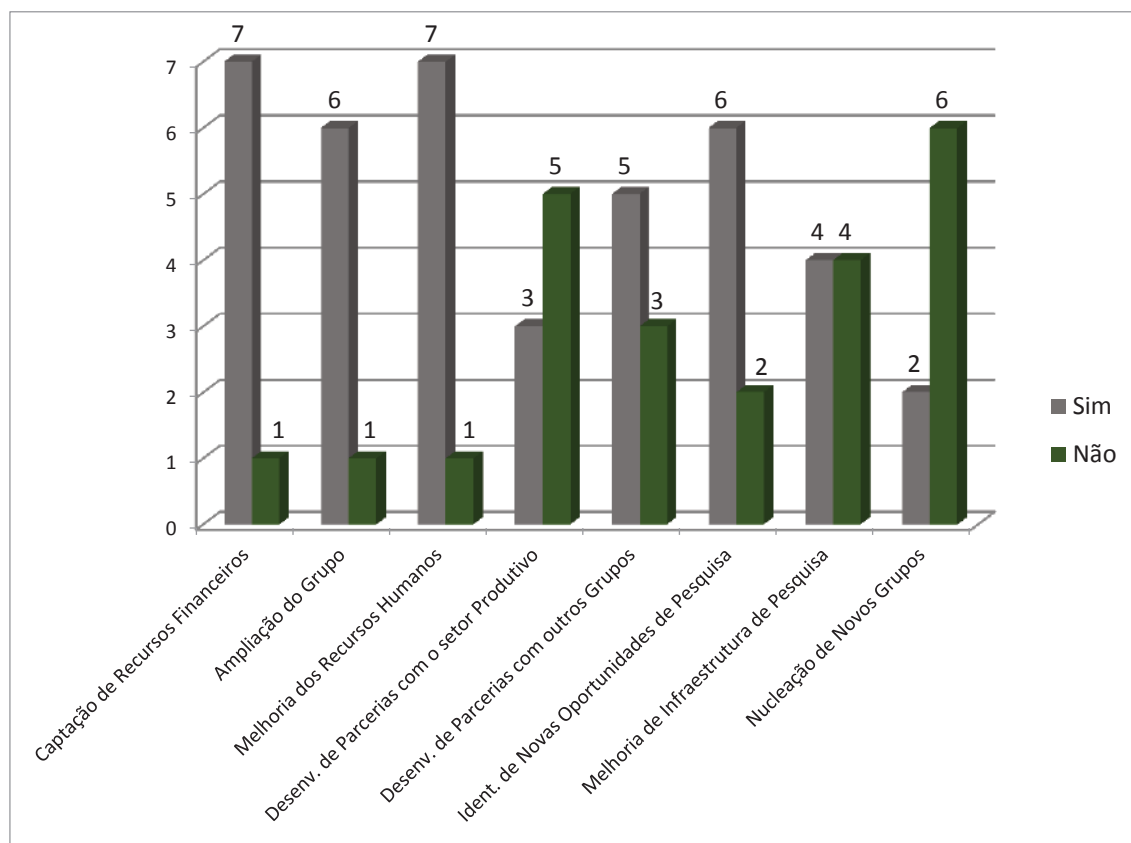


Figura 47. Objetivos de pesquisa dos pesquisadores do LES

Em uma mesma proporção 4-50% dos pesquisadores responderam que tem o objetivo de “melhoria de infraestrutura de pesquisa” e 4-50% dos pesquisadores responderam que não tem esse objetivo. A resposta negativa para este objetivo acontece quando os pesquisadores estão satisfeitos com a sua infraestrutura de pesquisa ou não é sua responsabilidade realizar esse tipo de melhoria.

5.1.7.2 Dados qualitativos das entrevistas

Na segunda etapa do questionário foram coletados dados qualitativos referentes às estratégias de pesquisas utilizadas pelos pesquisadores do LES. A **Pergunta**: “*Marque com ‘X’ os objetivos que você utiliza e relate as estratégias que você utiliza para alcançá-los*”, foi utilizada como base para esta coleta de dados. Para análise desses dados realizou-se uma síntese temática com o objetivo de apresentar os dados de forma sistemática, por ser utilizado em estudos de ES e possibilitar replicações do estudo e futuras comparações.

As etapas e recomendações descritas por Prates (2016) no Quadro 1 do Capítulo 2 foram adotados para analisar e extrair os dados qualitativos coletados das entrevistas. Essa abordagem foi adaptada para interpretar textos coletados de entrevistas visto que, os dados que serão analisados não serão dados de estudos primários. Desta forma, algumas recomendações foram descartadas durante o processo.

As etapas: extração dos dados, codificação de dados, tradução de códigos em temas, criação de um modelo de alto nível e avaliação da confiabilidade da síntese foram aplicadas para análise dos dados qualitativos das entrevistas dos pesquisadores e suas execuções estão descritas nas próximas subseções, juntamente com os resultados de cada uma. Apesar das etapas serem descritas sequencialmente no Quadro 1 e nas seções abaixo, as mesmas foram executadas de forma iterativa.

5.1.8 Extração dos dados

Para a preparação dos textos das entrevistas, após as suas transcrições, foram feitas cópias dos arquivos e as mesmas renomeadas. Os textos foram completamente relidos e durante esse processo concordâncias textuais foram corrigidas, frases, palavras duplicadas e textos que fossem considerados irrelevantes para os objetivos da pesquisa foram excluídos. Esta releitura se deu para que o autor desta dissertação imergisse e conhecesse melhor os

dados. Nos arquivos, segmentos específicos de texto (parágrafos, frases, palavras) referentes aos objetivos da síntese foram rotulados por meio de notas de comentários estabelecidas para cada marcação. Este processo foi realizado para todos os textos de entrevistas. A Figura 48 apresenta um exemplo das marcações feitas nos textos das entrevistas.

Entrevista com Manoel Gomes de Mendonça Neto (18/01/2017)

0:00 Começa a entrevista
Falo um pouco da pesquisa

03:13 Apresento termo de consentimento

04:30 O Entrevistado começa apreciar o material e o Entrevistador esclarece dúvidas.

12:50 O Entrevistado questiona o que é BTS (*Bug Tracking System*).

19:33 O Entrevistado relatou que as pessoas utilizam estudo de caso para estudo exploratório, o Entrevistador informou que utilizou para classificar nas categorias dos métodos de avaliação o que o autor disse no estudo. O Entrevistado pediu para colocar isso como um risco a validade a ser citado.

24:37 O Entrevistado pede para que as proporções dos gráficos de comparação com os pesquisadores (4.2 a 4.6) sejam esclarecidas junto aos pesquisadores para que eles entendam os gráficos.

26:52 O Entrevistado se surpreende que tem publicações que utilizam comentários em que ele não é coautor.

29:11 Início do questionário

29:17 1º etapa do questionário. Entrevistado responde aos questionamentos.

29:27 Resposta da 1ª pergunta: Eu acho que é uma das principais. Eu não diria que é a principal. A principal é visualização, mas, sempre combinada com muitas ideias de mineração.

29:57 Resposta da 2ª pergunta: O principal é o conhecimento que eu tenho em que a área engenharia tem que ser baseada em dados e hoje a área é muito intensiva em informação, ou seja, tem muitos dados disponíveis, o próprio código fonte é a maior fonte de dados, talvez por isso tem muita coisa feita pela gente minerando código fonte, mas também tem muitas coisas também que podemos minerar ainda. Então eu acho que é uma importante estratégia de pesquisa em engenharia de software.

30:49 Resposta da 3ª pergunta: Eu acho que sim. Tem algumas coisas que eu vi aqui. Deixa eu revisitar. Me permitiu ver nas minhas e no do grupo. Uma das coisas que eu achei importante que não tem nenhuma parte de mineração sobre o grupo especificamente em

[MS1] Comentário: RISCO A VALIDADE

[MS2] Comentário: SUGESTAO

[MS3] Comentário: SUGESTAO

[MS4] Comentário: COMENTARIO

[MS5] Comentário: AREA PRINCIPAL

[MS6] Comentário: AREA PRINCIPAL

[MS7] Comentário: OBJETIVO

[MS8] Comentário: MOTIVACAO

Figura 48. Exemplo de marcações feitas nos textos das entrevistas

Para essa etapa não foram utilizadas ferramentas para a extração dos dados, visto que o processo pode ser realizado de forma manual (Cruzes e Dyba 2011a; Prates 2016). Dessa forma, para auxiliar a extração dos dados foram utilizadas planilhas eletrônicas para registrar os dados extraídos e realizar a síntese temática. Essas planilhas serviram como instrumento de registro e como forma de tornar essa etapa mais confiável, o que diminui a possibilidade de retrabalhos.

Assim, após as marcações, os dados foram organizados em uma planilha eletrônica na aba Dados Gerais ou Dados das Estratégias, conforme o rótulo de cada marcação (Fig. 49).

Na primeira aba (Dados Gerais) foram armazenados todos os segmentos de texto que foram marcados com o objetivo de dar uma visão geral para responder aos objetivos que a entrevista se propôs, além de armazenar dados quantitativos que fossem relevantes para

responder algumas das perguntas feitas na primeira etapa dos questionários das entrevistas. Os rótulos de cada marcação foram relacionados aos objetivos da entrevista, como: avaliação do método, comentários, sugestões, dentre outros. Alguns rótulos não foram considerados na síntese (ex.: comentários), outros foram considerados na melhoria do material das entrevistas ou do método proposto (ex.: comentários, sugestões) e outros utilizados para entender o que os pesquisadores estavam pensando a respeito do método (ex.: avaliação do método). Então apesar da não utilização de todos os dados qualitativos na síntese, eles contribuíram para a interpretação e entendimento das ideias dos entrevistados. A Figura 49 apresenta a aba Dados Gerais na qual foram armazenados esses resultados.

	A	B	C	D	E	F
1	Entrevistado	Sugestões	Risco a Validade	Aprovação da metodologia	Dificuldades do Grupo	Comentários
2	C. Chavez	sugere criar uma tabela contendo as publicações por evento	questiona se é legal misturar os períodos, ou seja, 2010-2014 do MSR e 2010-2015 do LES. Pediu para ver melhor com Manoel, pois não acha interessante, porque acha que distorce os resultados	é sempre bom que tenha um trabalho que faça esse estudo, que confronte, que mostre visualmente, contrapor duas ou três dimensões, você acaba enxergando com mais clareza, caracterizando melhor	a gente não tem conseguido como LES traçar objetivos, metas para esse grande conjunto que é o LES	
3		na dissertação tem uma distinção do que foi considerado como MSR e pede para colocar no texto	A única coisa que eu não gostei foi incluir um ano a mais no LES	ainda mais com os mapas que ajudam a gente a entender onde podemos melhorar, onde podem ser oportunidades, onde estamos mais forte como grupo como pesquisador, acho interessante, e situar, está relacionando o pesquisador individualmente com o grupo e com os termos em cima de publicações e com a principal conferência na área		
4		questiona se foi produzido alguma coisa por instituição, pois com isso há uma percepção referente a quais instituições a UFPA interage mais		É legal porque a pessoa se situa, é interessante pro grupo, principalmente para quem está coordenando o grupo, porque vê para aquele tema específico, mas o grupo é muito mais, a gente tá vendo a pesquisa em relação a MSR, mas ficou legal		

Figura 49. Planilha de extração – aba Dados Gerais

Na segunda aba da planilha, Dados das Estratégias, foram armazenados os segmentos de texto referentes à segunda etapa do questionário das entrevistas relacionados às estratégias de pesquisas adotadas pelos pesquisadores (ex. estratégias, comentários). Todos os segmentos de textos extraídos para essa planilha que foram rotulados como estratégia e objetivo de pesquisa foram utilizados para realização da síntese temática. A Figura 50 apresenta a aba Dados das Estratégias na qual foram armazenados esses resultados.

	A	B	C	D	E	F	G	M
1	Entrevistado	Captação de Recursos Financeiros	Estratégias	Comentários	Ampliação do Grupo	Estratégias	Comentários	
3	J. SANTOS	Sim.	Através dos editais de bolsas e de projetos		Sim.	A minha estratégia hoje é entrar no programa de mestrado porque isso vai me habilitar a conseguir orientar alunos, com isso ter mais gente para formar e trabalhar nos editais de bolsas e de projetos que quero submeter, pois tendo recursos consigo mais gente		Sim.
4			é interessante fazer com que a comunidade local entenda a comunidade de se fazer pesquisa, mas ela não tem dinheiro para fazer pesquisa, então a gente tem que alinhar objetivos do setor produtivo com os objetivos específicos de em editais em particular e objetivos obviamente do grupo de pesquisa para fazer com que isso seja de interesse para a comunidade local e de dentro desses são mais interessantes a captação dos recursos humanos e ele tem uma relação direta com o desenvolvimento de pesquisas com o setor produtivo a melhoria da					

Figura 50. Planilha de extração – aba Dados das Estratégias

Todos os resultados desta etapa foram verificados por outro pesquisador, que validou as informações extraídas. Cada processo que foi realizado pode ser rastreado pelo nome do pesquisador, que concedeu a entrevista, nos arquivos e planilhas.

5.1.9 Codificação de Dados

Segundo Cruzes e Dyba (2011a) codificação significa nomear segmentos de textos com um rótulo que simultaneamente categoriza, resume e contabiliza cada segmento. Esta etapa é o primeiro passo para se realizar interpretações analíticas.

Após a etapa de extração dos dados outra aba foi criada na planilha eletrônica chamada Síntese Temática. Nela os segmentos já identificados e armazenados na aba Dados das Estratégias foram categorizados, codificados e classificados (Figura 51). O Quadro 9 apresenta a tabela com a descrição dos dados extraídos das entrevistas e armazenados na aba Síntese Temática da planilha de extração.

Quadro 9. Tabela dos dados extraídos das entrevistas (adaptado de Prates, 2016)

DADOS	DESCRIÇÃO
Pesquisador	Nome do entrevistado.
Objetivo	Objetivo de pesquisa do pesquisador
Texto	Segmento de texto completo da entrevista que contém evidências encontradas durante a leitura.
Resumo	Parágrafos, frases ou palavras que referencie as evidências do segmento de texto completo.
Código	Códigos que representam os itens contidos no resumo.
Classificação	Assunto principal do código.

	A	B	C	D	E	F
1	Entrevistado	Objetivo	Texto	Resumo	Código	Classificação
2	C. CHAVEZ	Captação de Recursos Financeiros	Via projetos, pensando tanto de infraestrutura, sempre editais, ou editais de pesquisa (fomentar pesquisa) como também parcerias com a indústria que eu não tenho feito recentemente mas e outra forma de captar recursos financeiro. É uma preocupação, mas não é algo que saímos planejando e tem estratégias para alcançar, isso acaba acontecendo	1- Via projetos, pensando tanto de infraestrutura, sempre editais, ou editais de pesquisa (fomentar pesquisa) como também 2 - parcerias com a indústria que eu não tenho feito recentemente mas é outra forma de captar recursos financeiro.	1-Editais de Pesquisa e Projetos [1] 2-Empresas Públicas e Privadas [2]	Estratégia para captação de Recursos Financeiros
3	J. SANTOS	Captação de Recursos Financeiros	Através dos editais de bolsas e de projetos	1 - Através dos editais de bolsas e de projetos	1-Editais de pesquisa e projetos [1]	Estratégia para captação de Recursos Financeiros
4	I. MACHADO	Captação de Recursos Financeiros	é interessante fazer com que a comunidade local entenda a comunidade de se fazer pesquisa, mas ela não tem dinheiro para fazer pesquisa, então a gente tem que alinhar objetivos do setor produtivo com os objetivos específicos de em editais em particular e objetivos obviamente do grupo de pesquisa para fazer com que isso seja de fato implementado na prática	1 - a gente tem que alinhar objetivos do setor produtivo com os objetivos específicos de em editais em particular e objetivos obviamente do grupo de pesquisa para fazer com que isso seja de fato implementado na prática	1-Editais de pesquisa e projetos [1]	Estratégia para captação de Recursos Financeiros

Figura 51. Planilha de extração – aba Síntese Temática

O processo de codificação pode ser realizado por meio de três abordagens:

- **Dedutiva:** uma lista de código é pré-definida através de teorias, hipóteses, etc., das áreas do problema e/ou das palavras-chave que o pesquisador traz ao estudo (Miles e Huberman 1994);
- **Indutiva:** os códigos, para um determinado conceito, surgem sem uma lista prévia (Corbin e Strauss 2007); e
- **Integrada:** é um meio termo entre a abordagem dedutiva e indutiva (Bogdan e Biklen 1982; Lofland 1971).

Nesta síntese utilizou-se a abordagem indutiva, visto que os segmentos de textos foram extraídos e um código foi atribuído observando-se o conceito mais aparente do segmento.

Este processo transformou dados qualitativos em classes de estratégias de pesquisa. Com ele as frequências de expressões para cada código e a quantidade de pesquisadores envolvidos em cada codificação puderam ser medidas. A consistência dos resultados foi verificada com a análise de outro pesquisador que realizou uma segunda análise para que os dados fossem confirmados ou corrigidos, o que melhorou a confiabilidade e credibilidade dos dados extraídos e da codificação.

5.1.10 Tradução de códigos em temas

Os temas segundo Saldaña (2008) não são de forma independentes codificados. Eles são resultado da codificação, categorização e reflexão analítica. O processo de tradução dos códigos em temas consiste em sintetizar diferentes códigos e criar um tema abrangente. Percebe-se com a Figura 52 que ao passo que os textos são resumidos, aumenta a abstração e a generalização da definição dos temas.

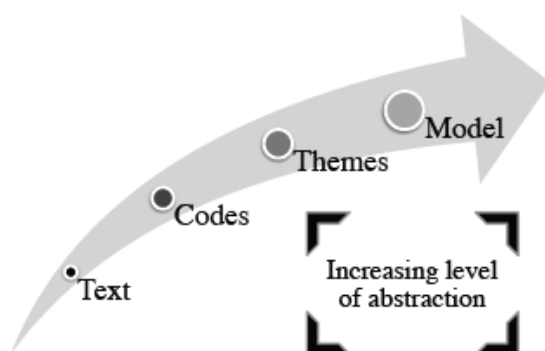


Figura 52. Nível de interpretação na síntese temática (Cruzes e Dyba, 2011a)

Como esta síntese temática foi realizada para identificar estratégias de pesquisas adotadas pelos pesquisadores do LES, automaticamente foi esperada a identificação de temas

ligados a tipos de estratégias de pesquisa. Desta forma, foram identificados os temas: (i) estratégias para captação de recursos financeiros, (ii) estratégias para melhoria dos recursos humanos, (iii) estratégias para ampliação do grupo, (iv) estratégias para desenvolvimento de parcerias com o setor produtivo, (v) estratégias para identificação de novas oportunidades de pesquisa, (vi) estratégias para melhoria da infraestrutura de pesquisa e (vii) estratégias para nucleação de novos grupos.

5.1.11 Criação de um modelo de alto nível

Nesta etapa todos os temas identificados na etapa anterior foram interpretados ao nível que temas de ordem superior fossem identificados e a relação entre eles apresentada. Segundo Cruzes e Dyba (2011a) e Prates (2016), como resultado desta etapa a produção de um mapa temático, uma taxonomia, uma visão mais descritiva dos temas ou uma teoria é realizada.

Desta forma, uma representação visual foi utilizada para melhor ilustrar os resultados obtidos. A Figura 53 apresenta o mapa temático produzido para representar a síntese temática.

Conforme a Figura 53 e descrito na seção anterior, sete temas ligados ao tema central “Estratégias de Pesquisa” foram identificados. As declarações feitas pelos pesquisadores nas entrevistas foram de fundamental importância para a realização da codificação e definição dos temas para atingir o resultado deste mapa temático que evidencia todas as estratégias de pesquisa utilizadas pelos pesquisadores do LES (temas) e de que forma elas são aplicadas (códigos).

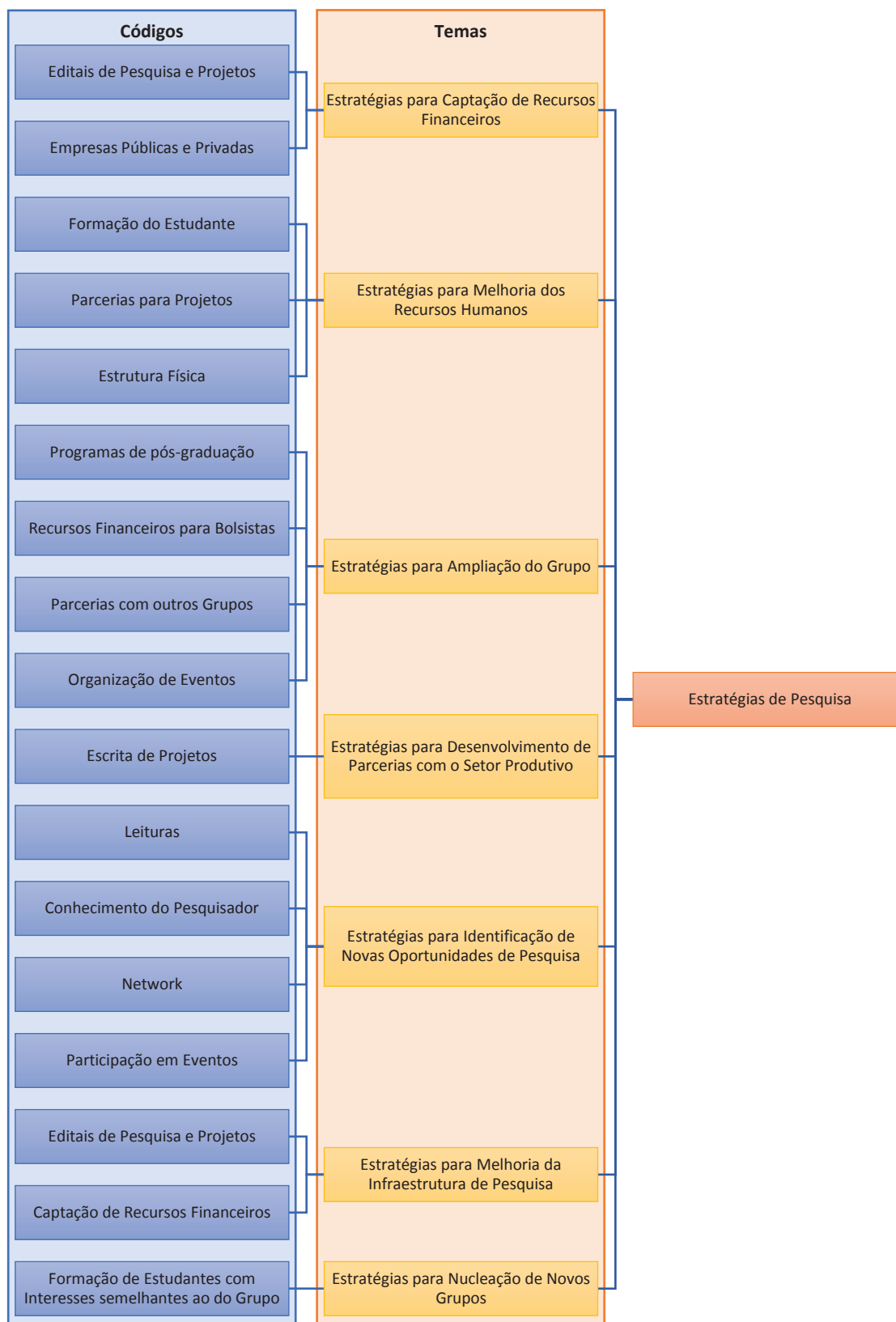


Figura 53. Mapa Temático para identificar e classificar estratégias de pesquisa

É objetivo deste trabalho criar um mapa temático para identificar e classificar as estratégias de pesquisa adotadas por grupos de pesquisa. O mapa temático da Figura 53 foi desenvolvido com essa finalidade e utilizou-se dos resultados colhidos durante a síntese temática realizada. Ele é artefato integrante do MASG e pode ser reutilizado em análises de outros grupos.

Com os resultados da síntese temática também foi identificado que o LES utiliza todas as estratégias de pesquisa apresentadas no mapa temático (Figura 53). Algumas com maior frequência como as “Estratégias para Captação de Recursos Humanos” (tema) por meio de “Editais de Pesquisa e Projetos” (código) que é utilizada por 87,5% dos pesquisadores do grupo e outras com menos frequência como “Estratégias para Identificação de Novas Oportunidades de Pesquisa” (tema) por meio de “Participação em Eventos” (código) que é utilizada por 12,5% dos pesquisadores do LES. A quantidade de expressões que justificaram os códigos e a quantidade de pesquisadores que mencionaram as expressões relacionadas a eles estão disponíveis no nosso repositório¹⁹.

5.2 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentado um estudo de caso que contempla a aplicação do Método para Análise Sistemática de Grupos de Pesquisa (MASG) proposto nesta dissertação (Figura 6). Percebe-se que os principais objetivos do método foram alcançados neste estudo, como a caracterização e posicionamento do LES em relação à área de MSR, o que possibilita o direcionamento dos líderes de pesquisa e pesquisadores do LES no desenvolvimento de novas pesquisas na área.

Durante a execução do MASG foram realizadas entrevistas com os pesquisadores do LES na busca de identificar as estratégias de pesquisa utilizadas pelo grupo, bem como, dados adicionais para a conclusão do estudo.

¹⁹ http://mestrado.web2405.uni5.net/Sintese_Tematica.xlsx

Essas entrevistas foram realizadas com os pesquisadores doutores do LES que tiveram publicações identificadas relacionadas com a área de MSR e apresentaram resultados quantitativos e qualitativos que evidenciaram a importância do método proposto, os tipos de mapas preferidos pelos pesquisadores e seus objetivos de pesquisa.

Também, proveniente dos dados qualitativos obtidos com a síntese temática executada com os textos das entrevistas, foi apresentado um mapa temático para identificação e classificação de estratégias de pesquisas. Este mapa permitiu identificar as estratégias de pesquisa que vem sendo utilizadas pelo LES e pode ser reutilizado por outros grupos de pesquisa que atuam com objetivos similares.

O próximo capítulo apresenta a discussão dos resultados e conclusão deste trabalho de dissertação, juntamente com as principais contribuições e limitações, ameaças à validade e trabalhos futuros que podem decorrer desta dissertação.

Neste capítulo são apresentadas as discussões dos resultados e conclusão desta dissertação juntamente com as suas contribuições, limitações e ameaças à validade. Sugestões de trabalhos futuros também são apresentadas.

6 CONCLUSÃO

6.1 DISCUSSÃO DOS RESULTADOS

Neste trabalho o Método de Análise Sistemática de Grupos de Pesquisa (MASG) foi proposto e devidamente descrito. Este método foi avaliado por meio de um estudo de caso aplicado no LES em publicações da área de MSR.

Durante a execução do método foram localizadas 367 publicações de 2010 a 2015 de 18 pesquisadores do LES, selecionados por meio de conversa com um especialista. Dessas publicações 59, relacionadas com a área de MSR, foram selecionadas para extração e análise de dados.

De forma sumarizada, percebeu-se que o LES tem a área de MSR como uma área de apoio para a maioria das suas pesquisas, por esse motivo muitas lacunas e oportunidades de pesquisa foram evidenciadas com este estudo. Com as comparações produzidas com a utilização do MASG, observou-se que ele pode auxiliar os líderes de pesquisa e pesquisadores a obterem o perfil do seu grupo frente a uma determinada área de pesquisa, o que possibilita futuras tomadas de decisão.

Foram realizadas entrevistas com oito pesquisadores do LES com o objetivo de identificar as estratégias de pesquisa utilizadas pelo grupo, bem como, colher dados que

pudessem trazer informações importantes para avaliação e melhoria do método. As estratégias de pesquisas identificadas durante esta atividade foram sistematicamente sumarizadas e deu origem a um mapa temático, uma das contribuições deste trabalho.

Como resultado geral alcançou-se o desenvolvimento e avaliação de um método para caracterização de publicações de grupo de pesquisa, bem como a identificação das estratégias de pesquisa utilizadas por ele, o que ajuda líderes de pesquisa e pesquisadores a alinharem as suas estratégias de pesquisa as tendências atuais, planejar e direcionar o desenvolvimento de novas pesquisas. Este dado foi evidenciado por meio das entrevistas e resultados alcançados durante a execução do MASG.

Apesar desta afirmação não poder ser generalizada, percebe-se por meio da atividade de entrevistas que os líderes de grupos e pesquisadores, muitas vezes, não têm uma visão do que está acontecendo no seu grupo de pesquisa em determinadas áreas. Trabalhos que utilizem o MASG podem dar essa visão.

A utilidade do MASG se justifica pela necessidade que os pesquisadores têm de se posicionar em relação a uma determinada área de interesse, geralmente vista por meio de mapeamentos sistemáticos, formais ou não. Atrelado a estes estudos, com o MASG o perfil de um grupo de pesquisa em relação a uma área de interesse pode ser traçado. O maior desafio dos pesquisadores é definir entre esses mapeamentos sistemáticos, quais são os melhores para conduzir o processo do MASG. Desta forma, essa atividade caracteriza-se como um trabalho futuro a ser realizado.

A utilização sistemática do MASG por diferentes grupos de pesquisa também possibilitará comparações sistemáticas entre grupos e pesquisadores da mesma ou de diferentes instituições.

6.2 RESULTADOS E CONTRIBUIÇÕES DESTE TRABALHO

As contribuições desta dissertação podem ser sumarizadas conforme descrito a seguir:

- Definição de um método para analisar trabalhos desenvolvidos por grupos de pesquisa, o que ajuda pesquisadores a alinharem as suas estratégias de pesquisas às tendências atuais e planejar novas pesquisas por meio de comparações entre mapas.
- Mapa visual, amplo e geral das comparações da área de MSR com as publicações relacionadas à mesma área do LES, bem como, uma taxonomia atualizada para pesquisas em MSR, o que possibilita novas comparações com diferentes grupos de pesquisa e pesquisadores.
- Definição de um mapa temático para identificar e classificar estratégias de pesquisa adotadas por um grupo.
- Realização de um estudo de caso, o que possibilita que outros grupos de pesquisa possam valer-se das informações sistematizadas como base de novas comparações e facilita o uso do método.
- Desenvolvimento de um roteiro de entrevistas para orientar a coleta de dados sobre as estratégias de pesquisa adotadas pelo grupo, bem como as opiniões dos pesquisadores quanto aos resultados obtidos com o método.

6.3 LIMITAÇÕES

Foram identificadas as seguintes limitações do projeto:

- O método abrange somente a parte de captura das estratégias usadas no presente e não as que podem ser usadas no futuro.
- O foco em MSR cobre apenas parte dos estudos do LES, devido à sua grande quantidade de publicações de áreas diversas.
- Para execução do método proposto, a classificação foi realizada baseada principalmente em pesquisadores que não são especialistas em Mineração de Repositório de Software. Apenas um pesquisador envolvido no projeto

tem alto conhecimento na área, apesar de não ser sua área principal de pesquisa.

- Realização de todos os processos de forma manual, sendo necessária muita energia para conseguir extrair informações de forma sistemática, analisar e apresentar resultados. A adoção ou desenvolvimento de uma ferramenta facilitaria a tarefa de extração e análise dos dados e dos textos das entrevistas, com a diminuição do trabalho e a complexidade inerente.
- Para que o método seja executado, se faz necessária a existência de um MS na área de interesse.

6.4 AMEAÇAS À VALIDADE

A validade de um experimento está relacionada ao nível de confiança do processo executado de investigação experimental (Travassos *et al.*, 2002). Segundo Wohlin *et al.* (2012) as ameaças à validade dos resultados de estudos experimentais podem ser: interna, externa, de construção e de conclusão, mas, nem todas são aplicáveis a todos experimentos. Segue abaixo os principais riscos à validade identificados e as ações realizadas na tentativa de mitigá-los.

Validade Interna. A área temática escolhida para realizar o estudo de caso desta dissertação foi Mineração de Repositório de Software (MSR). Acredita-se que a escolha da área influenciou nas entrevistas, pois os pesquisadores que possuem essa área como principal se mostraram mais interessados pelos resultados, enquanto os outros, em alguns casos, mostram desinteresse.

Validade de Construção. De início a primeira seleção das publicações do LES realizada durante a **Atividade 3 – Selecionar as publicações do grupo** – do MASG foram feitas somente pelo autor desta dissertação. Isto significa que alguns estudos podem ter ficado de fora da classificação. Com a intenção de mitigar esta ameaça após a seleção foram

realizadas conversas com alguns pesquisadores do LES com o objetivo de verificar se alguma publicação tinha ficado fora da seleção.

No processo de classificação dos dados das publicações, as mesmas podem ter sido classificadas incorretamente. Com a intenção de mitigar esta ameaça, dois pesquisadores leram cada publicação e um terceiro pesquisador participou da discussão quando necessário.

Utilizou-se para execução do método publicações de 2010 a 2014 da área de MSR e de 2010 a 2015, ou seja, um ano a mais do LES, o que possibilita o surgimento de assuntos de um ano para o outro que podem não está capturado na área e está capturado no LES, o que pode distorcer os resultados. Este risco foi assumido para que pudéssemos ter uma densidade maior de publicações do LES a serem analisadas.

Utilizou-se o que o autor disse no texto das publicações do LES, como parâmetro de classificação para a categoria “métodos de avaliação”. Este critério também foi utilizado pelo mapeamento sistemático da área de MSR, o que pode ter gerado classificações incorretas. Para mitigar esse risco os pesquisadores, caso houvesse dúvidas, a classificação dos métodos de avaliação eram confirmadas por meio de leitura crítica do processo de avaliação descrito no estudo.

Validade Externa. O método proposto sugere a utilização da fonte de dados Plataforma Lattes para identificar as publicações do grupo por considerar que a mesma se encontra sempre atualizada, por ser fonte de avaliação dos pesquisadores no Brasil. A desatualização, ou mesmo indisponibilidade do Lattes, no caso de pesquisadores de fora do Brasil, poderia inviabilizar a aplicação do método, se uma fonte padrão de CV não estiver disponível. No nosso estudo de caso, para evitar que essa fonte de dados, para alguns pesquisadores, estivesse desatualizada, foram feitas verificações manuais e pedido aos pesquisadores que seus currículos lattes fossem atualizados, pois seriam utilizados na análise.

A classificação feita pelo ScriptLattes segue as informações de registro feitas pelo pesquisador ao cadastrar o seu currículo lattes. Então isso pode causar um ruído de informação, pois os pesquisadores classificam suas publicações de formas diferentes, isso se dá pela inexperiência dos pesquisadores com a Plataforma Lattes ou definições de modelo e registro a seguir (instruções de orientadores, sugestões de pesquisadores, etc.). Com o objetivo de diminuir esses ruídos, alguns erros foram informados aos pesquisadores durante a primeira execução do ScriptLattes, desta forma, alguns fizeram as correções apontadas e

outros não foram localizados durante o prazo estabelecido para as modificações. Após este prazo o ScriptLattes foi executado novamente. Alguns casos foram identificados quando os arquivos PDFs das publicações foram recuperados. Percebeu-se que alguns registros não eram publicações, e sim, banners, mesas redondas, etc. e outros foram classificados pelos ScriptLattes como “não identificados” e mesmo assim somavam quantitativamente em seus relatórios, mas foram descartados durante o processo.

6.5 TRABALHOS FUTUROS

Como trabalhos futuros podem ser realizadas atividades para refinar e aperfeiçoar a aplicação do método proposto nesta dissertação, como:

- Definir os critérios para identificar quais os melhores MS para executar o MASG.
- Realizar novas aplicações práticas do MASG no LES, visto que, o laboratório possui diversos MS, em diferentes áreas de interesse, já conduzidos, que podem ser utilizados (ex. Alves *et al.* 2016; Dusse *et al.*, 2016; Santos *et al.*, 2015).
- Realizar aplicações práticas do MASG em outros grupos e fazer comparações entre elas;
- Desenvolver uma ferramenta que automatize o método proposto ou algumas atividades como a de comparação dos mapas.
- Sumarizar os resultados obtidos e escrever artigos científicos para conferências da área.

REFERÊNCIAS

- AFZAL, W.; TORKAR, R.; FELDT, R. A Systematic Mapping Study on Non-Functional Search-based Software Testing. In: **SEKE**. p. 488-493, 2008.
- ALMEIDA, A.; BARREIROS, E.; SARAIVA, J.; SOARES, S. Mecanismos para Guiar Estudos Empíricos em Engenharia de Software. Um mapeamento sistemático. In: **Proceedings of 8 th Experimental Software Engineering Latin American Workshop**. 2011. p. 37.
- ALVES, N. S.; MENDES, T. S.; DE MENDONÇA, M. G.; SPINOLA, R. O.; SHULL, F.; SEAMAN, C. Identification and management of technical debt: A systematic mapping study. In: **Information and Software Technology** 70. p. 100-121, 2016.
- ARKSEY, H., O'MALLEY, L. Scoping studies: towards a methodological framework. **International Journal of Social Research Methodology**, v. 8, n. 1, p. 19-32, 2005.
- BAILEY, J.; et. al., Evidence relating to Object-Oriented software design: A survey. In: **Symposium on Empirical Software Engineering and Measurement**. p. 482-484. 2007.
- BALANCIERI, R.; BOVO, A. B.; KERN, V. M.; PACHECO, R. C. S.; BARCIA, R. M. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**. 2005. 34(1): 64-77.
- BASIL, V. R., The past, present, and future of experimental software engineering. **J. Braz. Comp. Soc.**, vol.12, n.3, pp.7-12, 2006.
- BASIL, V. R., SHULL, F., LANUBILE, F. Building knowledge through families of experiments. **Software Engineering, IEEE Transactions on**, v. 25, n. 4, p. 456-473, 1999.
- BASIL, V. R.; WEISS, David M. A methodology for collecting valid software engineering data. **Software Engineering, IEEE Transactions on**, n. 6, p. 728-738, 1984.
- BERRY, M. J., LINOFF. G. **Data mining techniques: for marketing, sales, and customer support**. John Wiley & Sons, Inc., 1997.
- BOGDAN, R.; BIKLEN, S.K. Qualitative research for education: An introduction to theory and methods. 2d ed. Boston: Allyn & Bacon, 1982.
- BRADLEY, E. H.; CURRY, L. A.; DEVERS, K. J. Qualitative Data Analysis for Health Services Research: Developing Taxonomy, Themes, and Theory. **Health Services Research**, 42(4):1758–1772, August 2007.
- BIOLCHINI, J. C. A., *et al.*, Scientific research ontology to support systematic review in software engineering. **Advanced Engineering Informatics**, v. 21, n. 2, p. 133-151, 2007.

BIOLCHINI, J.; TRAVASSOS, G. H. Revisões sistemáticas aplicadas a engenharia de software. In **XXI SBES - Brazilian Symposium on Software Engineering**, João Pessoa, PB, Brasil, 2007a.

BIOLCHINI, J.; *et al.*, Systematic review in software engineering. **System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES**, v. 679, n. 05, p. 45, 2005.

BRAUN, V.; CLARKE, V. Using thematic analysis in psychology. **Qualitative Research in Psychology**, 3(2):77-101. Routledge, 2006.

BRERETON, P.; *et al.*, Lessons from applying the systematic literature review process within the software engineering domain. **Journal of systems and software**, v. 80, n. 4, p. 571-583. 2007.

BUDGEN, D., *et al.*, Using Mapping Studies in Software Engineering. In: **Proceedings of PPIG**, p. 195-204, 2008.

CAMPBELL, R. *et al.* Evaluating meta-ethnography: a synthesis of qualitative research onlay experiences of diabetes and diabetes care. **Soc Sci Med**, 2003. 671-684.

CAVALCANTI, Y. C.; SILVEIRA NETO, P. A. M.; LUCREDIO, D.; VALE, T.; ALMEID, E. S. A.; MEIRA, S. R. L. The bug report duplication problem: an exploratory study. **Software Quality Journal**, v. 21, no. 1. p. 39-66. 2013.

CHIZZOTTI, Antonio. **Pesquisa qualitativa em ciências humanas e sociais**. Vozes, 2006.

COHEN, Aaron M. *et al.*, Reducing workload in systematic review preparation using automated citation classification. **Journal of the American Medical Informatics Association**, v. 13, n. 2, p. 206-219, 2006.

COUNSELL, C. Formulating questions and locating primary studies for inclusion in systematic reviews. **Ann Intern Med**, Setembro 1997. 380-719.

CONDORI-FERNANDEZ, N., *et al.*, A systematic mapping study on empirical evaluation of software requirements specifications techniques. In: **Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement**. IEEE Computer Society. p. 502-505, 2009.

CONDORI-FERNANDEZ, N.; DANEYA, M.; SIKKEL, K.; WIERINGA, R.; DIESTE, O.; PASTOR, O. A systematic mapping study on empirical evaluation of software requirements specifications techniques. In: **Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement**. IEEE Computer Society. p. 502-505, 2009.

- COOK, T. D.; REICHARDT, C. S. (Eds.). **Qualitative and quantitative methods in evaluation research**. Vol. 1. Beverly Hills, CA: Sage publications, 1979.
- CORBIN, J.; STRAUSS, A. **Basics of Qualitative Research** (third ed.), Sage, 2007.
- CRESWELL, J. W. Projeto de pesquisa métodos qualitativo, quantitativo e misto. In: **Projeto de pesquisa métodos qualitativo, quantitativo e misto**. Artmed, 2010.
- CRESWELL, John W.; CLARK, Vicki L. Plano. **Designing and conducting mixed methods research**. 2007.
- CRUZES, D. S.; DYBA, T. Recommended Steps for Thematic Synthesis in Software Engineering. In: **Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement**, 2011a. 275-284.
- CRUZES, D. S.; DYBA, T. Research synthesis in software engineering: A tertiary study. **Information and Software Technology**, v. 53, n. 5, p. 440-455, 2011b.
- DA SILVA, Fabio QB *et al.*, Six years of systematic literature reviews in software engineering: An updated tertiary study. **Information and Software Technology**. v. 53, n. 9, p. 899-913, 2011.
- DEMEYER, S.; MURGIA, A.; WYCKAMANS, K.; LAMKANFI, A. Happy birthday! A trend analysis on past MSR papers. **Int. Work. Conf. Min. Softw. Repos.** pp. 353-362, 2013.
- DUARTE, Rosália. Entrevistas em pesquisas qualitativas Interviews in qualitative research. **Educar em revista**, v. 24, p. 213-225, 2004.
- DUSSE, F.; JÚNIOR, P. S.; ALVES, A. T.; NOVAIS, R., Vieira, V.; MENDONÇA, M. Information visualization for emergency management: A systematic mapping study. In: **Expert Systems with Applications** **45**. p. 424-437, 2016.
- DYBA, T., DINGSØYR, T., HANSEN, G. K. Applying systematic reviews to Diverse Study Types: An Experience Report. In: **Proceedings of the First International Symposium on Empirical Software Engineering and Measurement**, p. 225-234, 2007.
- DYBA, T.; DINGSØYR, T. Strength of evidence in systematic reviews in software engineering. In: **Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement**. ACM, p. 178-187, 2008.
- DYBA, T.; KITCHENHAM, B. A.; JORGENSEN, M. Evidence-based software engineering for practitioners. **IEEE software**, v. 22, n. 1, p. 58-65. 2005.
- EASTERBROOK, S., *et al.*, Selecting empirical methods for software engineering research. In: **Guide to advanced empirical software engineering**. Springer London. p. 285-311, 2008.

FARIAS, M. A. F.; MENDONÇA NETO, M. G.; DA SILVA, A. B.; SPÍNOLA, R. O. A Contextualized Vocabulary Model for identifying technical debt on code comments. In **Managing Technical Debt (MTD), 2015 IEEE 7th International Workshop on**. pp. 25-32. IEEE, 2015a.

FARIAS, M. A. F.; NOVAIS, R. L.; ORTINS, P.; COLAÇO JR, M.; MENDONÇA, M. G. Analyzing Distributions of Emails and Commits from OSS Contributors through Mining Software Repositories-An Exploratory Study. In: **ICEIS (2)**, p. 303-310. 2015.

FARIAS, M.; NOVAIS, R.; JÚNIOR, C.; CARVALHO, L.; MENDONÇA, M.; SPÍNOLA, R. A systematic mapping study on mining software repositories. In: **Proceedings of the 31st Annual ACM Symposium on Applied Computing**. 2016. p. 1472-1479.

GIBBS, Graham. **Análise de dados qualitativos: coleção pesquisa qualitativa**. Bookman Editora, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo, v. 5, p. 61, 2002.

GODOY, Arlida Schmidt. Introdução à pesquisa qualitativa e suas possibilidades. **Revista de administração de empresas**, v. 35, n. 2, p. 57-63, 1995.

GÜNTHER, Hartmut. Pesquisa qualitativa versus pesquisa quantitativa: esta é a questão. **Psicologia: teoria e pesquisa**, v. 22, n. 2, p. 201-210, 2006.

HAMER, S.; GILL, C. **Achieving evidence-based practice: a handbook for practitioners**. 2^o Edição. ed. Londres: Baillière Tindall, 2005.

HEMMATI, S. N. H.; BAYSAL, O.; KONONENKO, O.; WANG, W.; HOLMES, R.; GODFREY, M. W. The MSR Cookbook: Mining a decade of research. **10th Work. Conf. Min. Softw. Repos**. May 2013. p. 343–352.

JORGENSEN, M., SHEPPERD, M. A systematic review of software development cost estimation studies. **Software Engineering, IEEE Transactions on**, v. 33, n. 1, p. 33-53, 2007.

JURGEN, M.; *et al.*, Software process definition and management. **Springer Science & Business Media**. 2012. Chapter 4.

KAGDI, H.; COLLARD, M. L.; MALETIC, J. I. A survey and taxonomy of approaches for mining software repositories in the context of software evolution. **Journal of Software Maintenance and Evolution: Research and Practice**, v. 19, n. 2, p. 77-131, 2007.

KHAN, Siffat Ullah; NIAZI, Mahmood; AHMAD, Rashid. Barriers in the selection of offshore software development outsourcing vendors: An exploratory study using a systematic literature review. **Information and Software Technology**, v. 53, n. 7, p. 693-706, 2011.

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

KITCHENHAM, B. A.; MENDES, E.; TRAVASSOS, G. H. Cross versus within-company cost estimation studies: A systematic review. **IEEE Transactions on Software Engineering**, v. 33, n. 5, p. 316-329. 2007a.

KITCHENHAM, B. A.; DYBA, T., JORGENSEN, M. Evidence-based software engineering. **Proceedings of the 26th international conference on software engineering**. IEEE Computer Society. p. 273-281, 2004.

KITCHENHAM, B. What's up with software metrics?—A preliminary mapping study. **Journal of systems and software**, v. 83, n. 1, p. 37-51, 2010.

KITCHENHAM, B., MENDES, E., TRAVASSOS, G. H. Systematic Review of Cross versus within-company cost estimation studies. In: **Proceedings of the 10th international conference on Evaluation and Assessment in Software Engineering**. p. 81-90, 2006.

KITCHENHAM, B.; BRERETON, O. P.; BUDGEN, D.; TURNER, M.; BAILEY, J.; LINKMAN, S. Systematic literature reviews in software engineering—a systematic literature review. **Information and software technology**. v. 51, n. 1, p. 7-15, 2009.

KITCHENHAM, B.; *et al.*, Systematic literature reviews in software engineering—a tertiary study. **Information and Software Technology**, v. 52, n. 8. p. 792-805. 2010.

KITCHENHAM, Barbara A.; BUDGEN, David; BRERETON, O. Pearl. Using mapping studies as the basis for further research—a participant-observer case study. **Information and Software Technology**, v. 53, n. 6, p. 638-651, 2011.

LAKATOS, E. M.; MARCONI, M. d. A. Fundamentos da metodologia científica. In: **Fundamentos da metodologia científica**. [S.l.]: Atlas. 2010.

LOFLAND, J.; Analyzing Social Settings: A Guide to Qualitative Observation and Analysis. Belmont, CA: **Wadsworth**, 1971.

LÜDERS, F. Use of Component-Based Software Architectures in Industrial Control Systems. **Department of Computer Science and Engineering**. Mälardalen University. Västerås, Sweden: Mälardalen University Press, 2003.

MAFRA, S. N., TRAVASSOS, G. H. Estudos Primários e Secundários Apoiando a Busca por Evidência em Engenharia de Software, **Relatório Técnico: RT-ES-687/06-Programa de Engenharia de Sistemas e Computação-COPPE/UFRJ-Rio de Janeiro**, 2006.

MENA-CHALCO, J. P.; JÚNIOR, C.; MARCONDES, R. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, 15(4). p.31-39. 2009.

MILES, M.B; HUBERMAN, A.M. Qualitative data analysis: An expanded sourcebook (2nd ed.). Thousand Oaks, CA: **Sage**, 1994.

MINAYO, M. C. S. O desafio do conhecimento. Pesquisa qualitativa em saúde. 2008.

NETO, P. A. D. M. S.; MACHADO, I. C. M.; MCGREGOR, J. D; ALMEIDA, E. S.; MEIRA, S. R. M. A systematic mapping study of software product lines testing. **Information and Software Technology**, v. 53, n. 5, p. 407-423. 2011.

NOVAIS, R. L., *et al.*, Software evolution visualization: A systematic mapping study. **Information and Software Technology**, v. 55, n. 11, p. 1860-1883, 2013.

NOVAIS, R. L.; TORRES, A.; MENDES, T. S.; MENDONÇA, M.; ZAZWORKA, N. Software evolution visualization: A systematic mapping study. **Information and Software Technology**, v. 55, n. 11, p. 1860-1883, 2013.

NSF, Chapter 5. Academic Research and Development Science and Engineering, **National Board Indicators 2016 Report. A broad base of quantitative information on the U.S. and international science and engineering enterprise.** Disponível em <<https://www.nsf.gov/statistics/2016/nsb20161/#/>>, acessado em 07 de março de 2017>

NUNES, E. D. O desafio do conhecimento: pesquisa qualitativa em saúde. **Ciência & Saúde Coletiva**, v. 12, n. 4, p. 1087-1088. 2007.

OATES, B. J.; CAPPER, G. Using systematic reviews and evidence-based software engineering with masters students. In: **13th International Conference on Evaluation and Assessment in Software Engineering**. p. 20-21, 2009.

PAI, M.; *et al.*, Systematic reviews and meta-analyses: an illustrated, step-by-step guide. **The National medical journal of India**, v. 17, n. 2, p. 86-95, 2003.

PATERSON, B. L. *et al.* **Meta-Study of Qualitative Health Research: A Practical Guide to Meta-Analysis and Meta-Synthesis.** Thousand Oaks, California: Sage Publications, Inc., v. III, 2001.

PERRY, D. E.; PORTER, A. A.; VOTTA, L. G. Empirical studies of software engineering: a roadmap. In: **Proceedings of the Conference on The Future of Software Engineering**, ACM, p. 345–355, 2000.

PETERSEN, K.; *et al.*, Systematic mapping studies in software engineering. In: **12th International Conference on Evaluation and Assessment in Software Engineering**. 2008. p. 71-80.

PETTICREW, Mark; ROBERTS, Helen. **Systematic reviews in the social sciences: A practical guide**. John Wiley & Sons, 2008.

PRATES, Luciana Carla Lins. **Aplicando síntese temática em engenharia de software**. 2016.

RICHARDSON, R. J. **Pesquisa Social - Métodos e Técnicas**. 3ª edição. ed. São Paulo: Atlas, 2008.

SALDAÑA, J. **The Coding Manual for Qualitative Researchers**. London: Sage, 2008.

SANTOS, A.R.; DE OLIVEIRA, R.P; DE ALMEIDA, E.S. Strategies for consistency checking on software product lines: a mapping study. In: **Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering**. p. 5, ACM, 2015.

SHAH, Sonali K.; CORLEY, Kevin G. Building better theory by bridging the quantitative–qualitative divide. **Journal of Management Studies**, v. 43, n. 8, p. 1821-1835, 2006.

THOMAS, J.; HARDEN, A. Methods for the thematic synthesis of qualitative research in systematic reviews. **BMC Med Res Methodol**, v. 8, n. 45, 2008.

TRAVASSOS, G. H.; BIOLCHINI, J. Revisões sistemáticas aplicadas a engenharia de software. In: **XXI SBES-Brazilian Symposium on Software Engineering**. 2007.

WOHLIN, C.; RUNESON, P.; HOST, M.; OHLSSON, M.; REGNELL, B.; WESSLEN, A. **Experimentation in Software Engineering: an Introduction**. Massachusetts: Kluwer Academic Publishers, 0-7923-8682-5, 2000.

WOHLIN, C.; RUNESON, P.; HOST, M.; OHLSSON, M. C.; REGNELL, B. & WESSLEN, A. **Experimentation in Software Engineering: An Introduction**. Springer, Norwell, MA, USA, 2012.

WEBSTER, J.; WATSON, R. T. Analyzing the past to prepare for the future: Writing a literature review. **MIS quarterly**, p. xiii-xxiii, 2002.

YAMASHITA, A.; COUNSELL, S. Code smells as system-level indicators of maintainability: An empirical study. **Journal of Systems and Software** **86 (10)**, 2013. 2639-2653.

APÊNDICE A

LISTA DOS PESQUISADORES ANALISADOS DO LES

ID	NOME DO PESQUISADOR	GRUPO	ABREVIÇÃO
1	Adolfo Almeida Duran	MEFES	A. Duran
2	Christina von Flach Garcia Chaves	Aside	C. Chavez
3	Cláudio Nogueira Sant'Anna	Aside	C. Santanna
4	Eduardo Santana de Almeida	RISE	E. Almeida
5	Ivan do Carmo Machado	RISE	I. Machado
6	Manoel Gomes de Mendonça	SoftVis	M. Mendonça
7	Rodrigo Rocha Gomes e Souza	Aside	R. Souza
8	Roberto Almeida Bittencourt	Aside	R. Bittencourt
9	José Amâncio Macedo Santos	SoftVis	J. Santos
10	Renato Lima Novais	SoftVis	R. Novais
11	Antônio Soares de Azevedo Terceiro	Aside	A. Terceiro
12	Bruno Carreiro da Silva	Aside	B. Silva
13	Gecynalva Soares da Silva Gomes	RISE	G. Gomes
14	Ivonei Freitas da Silva	RISE	I. Silva
15	Luanna Lopes Lobato	RISE	L. Lobato
16	Raphael Pereira de Oliveira	RISE	R. Silva
17	Yguaratã Cerqueira Cavalcanti	RISE	Y. Cavalcanti
18	Vaninha Vieira dos Santos	CEmanTIKA	V. Santos

APÊNDICE B

LISTA COMPLETA DOS PESQUISADORES QUE TIVERAM PUBLICAÇÕES IDENTIFICADAS NO ESTUDO

ID	NOME DO PESQUISADOR	ABREVIÇÃO
1	DE MENDONÇA, MANOEL G.	M. Mendonca
2	NOVAIS, R. L.	R. Novais
3	CHAVEZ, CHRISTINA	C. Chavez
4	ALMEIDA, E. S.	E. Almeida
5	SANT ANNA, CLAUDIO	C. Santanna
6	BITTENCOURT, R. ; A.	R. Bittencourt
7	GARCIA, A.	A. Garcia
8	MEIRA, SILVIO ROMERO DE LEMOS	S. Meira
9	CARNEIRO, G. F.	G. Carneiro
10	CAVALCANTI, Y. C.	Y. Cavalcanti
11	FARIAS, M. A. F.	M. Farias
12	SOUZA, R.	R. Souza
13	COLACO JUNIOR, M.	M. Jr
14	LIMA, C. A. N.	C. Lima
15	FIGUEIREDO, E.	E. Figueiredo
16	NETO, P. A. M. S.	P. Neto
17	SANTOS, JOSÉ AMÂNCIO M.	J. Santos
18	SIMÕES JR, PAULO R. M.	P. Jr
19	TERCEIRO, ANTONIO SOARES DE AZEVEDO	A. Terceiro
20	LUCRÉDIO, DANIEL	D. Lucredio
21	MENDES, T. S.	T. Mendes
22	PADILHA, J.	J. Padilha
23	BERTRAN, I. M.	I. Bertran
24	CUNHA, C. E. A.	C. Cunha
25	FIGUEIREDO, J. C. A.	J. Figueiredo
26	HENRIQUE, PAULO	P. Henrique
27	KON , FABIO.	F. Kon
28	MACHADO, I. C.	I. Machado
29	MEIRELLES, PAULO	P. Meirelles
30	MIRANDA, J.	J. Miranda
31	NUNES, CAMILA	C. Nunes
32	O. SPÍNOLA, RODRIGO	R. Spínola
33	ORTINS, P.	P. Ortins
34	RIOS, LUIZ ROMÁRIO	L. Rios
35	SANTOS, ALCEMIR	A. Santos

36	STAA, ARNDT VON	A. Staa
37	VALE, T.	T. Vale
38	A. ALMEIDA, DANIEL	D. Almeida
39	ALMEIDA, LUCIANNA	L. Almeida
40	ALMEIIDA, J.	J. Almeida
41	ANDRADE, S. S.	S. Andrade
42	ARCOVERDE, R.	R. Arcoverde
43	BATISTA, E.	E. Batista
44	BRUNET, JOAO	J. Brunet
45	CARVALHO, LUIS PAULO DA S.	L. Carvalho
46	CIRILO, ELDER	E. Cirilo
47	CORUMBÁ, DANIELA	D. Carumba
48	COSTA, JOENIO	J. Costa
49	CRNKOVIC, IVICA	I. Crnkovic
50	CRUZES, DANIELA SOARES	D. Cruzes
51	DAMASCENO, A.	A. Damasceno
52	DANTAS, FRANCISCO	F. Dantas
53	DE ANDRADE, HUGO SICA	H. Andrade
54	GUERRERO, D. D. S.	D. Guerrero
55	MAGNAVITA, R.	R. Magnavita
56	MAIA, M. C. O.	M. Maia
57	MARA, LEANDRA	L. Mara
58	MARON, DIOGO L.	D. Maron
59	MEDEIROS, T. F. L.	T. Medeiros
60	PAIVA, T.	T. Paiva
61	PEREIRA, J.	J. Pereira
62	PINHO, D.	D. Pinho
63	RAIMUNDO, P. O.	P. Raimundo
64	ROCHA, J. F.	J. Rocha
65	S. R. ALVES, NICOLLI	N. Alves
66	SANTOS JR., CARLOS	C. Jr
67	SANTOS, CLEBER P. DOS	C. Santos
68	SANTOS, FRANCISCO R.	F. Santos
69	SEREY, DALTON	D. Serey
70	SILVA, ANDRÉ BATISTA	A. B. Silva
71	SILVA, ARLESON N.	A. Silva
72	SILVA, BRUNO	B. Silva
73	SILVA, CARLOS V. A.	C. Silva
74	SILVA, D. S.	D. Silva
75	SILVA, MARCOS	M. Silva
76	SOUZA, I.	I. Souza
77	TEREZA, L.	L. Tereza

ANEXO 1

A SYSTEMATIC MAPPING STUDY ON MINING SOFTWARE REPOSITORIES

A Systematic Mapping Study on Mining Software Repositories

Mário André de F. Farias
Federal Institute of Sergipe and Federal University of Bahia
mario.andre@ifs.edu.br

Methanias Colaço Júnior
Federal University of Sergipe
mjrse@hotmail.com

Manoel Mendonça
Federal University of Bahia
manoel.mendonca@ufba.br

Renato Novais
Fraunhofer Proj.Center at UFBA and Federal Institute of Bahia
renato@ifba.edu.br

Luis Paulo da Silva Carvalho
Federal Institute of Bahia
luiscarvalho@ifba.edu.br

Rodrigo Oliveira Spinola
Fraunhofer Proj. Center at UFBA and Salvador University
rodrigo.spinola@pro.unifacs.br

ABSTRACT

Background: Software repositories provide large amount of data encompassing software changes throughout its evolution. Those repositories can be effectively used to extract and analyze pertinent information and derive conclusions related to the software history or its current snapshot. *Objective:* This work aims to investigate recent studies on Mining Software Repositories (MSR) approaches collecting evidences about software analysis goals (purpose, focus, and object of analysis), data sources, evaluation methods, tools, and how the area is evolving. *Method:* A systematic mapping study was performed to identify and analyze research on mining software repositories by analyzing five editions of Working Conference on Mining Software Repositories – the main conference on this area. *Results:* MSR approaches have been used for many different goals, mainly for comprehension of defects, analysis of the contribution and behavior of developers, and software evolution comprehension. Besides, some gaps were identified with respect to their goals, focus, and data source type (e.g. lack of usage of comments to identify smells, refactoring, and issues of software quality). Regarding the evaluation method, our analysis pointed out to an extensive usage of some types of empirical evaluation. *Conclusion:* Studies of the MSR have focused on different goals, however there are still many research opportunities to be explored and issues associated with MSR that should be considered.

CCS Concepts

• *General and reference—Surveys and overviews* • *Software and its engineering—Software configuration management and*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

SAC 2016, April 04–08, 2016, Pisa, Italy
© 2016 ACM. ISBN 978-1-4503-3739-7/16/04...\$15.00
DOI: <http://dx.doi.org/10.1145/2851613.2851786>

version control systems • *Software and its engineering—Software libraries and repositories*

Keywords

Systematic mapping study; mining software repository; empirical software engineering; secondary study.

1. INTRODUCTION

The Mining Software Repositories (MSR) field focuses on uncovering interesting and useful information about software and projects extracting and analyzing available data from different software repositories [1].

Software repositories contain large amount of software historical data that can include valuable information on the source code, defects, and other issues like new features. Thereby, MSR have become an important area for Software Engineering (SE) researches and a popular tool for empirical studies in this area [2]. Several researches have been exploring MSR, aiming to discuss topics such as prediction of defects [3][4][5] and comprehension of software evolution [6][7][8].

In this context, since 2004 an annual workshop called Working Conference on MSR (MSRConf) has been a primary venue for publishing and discussing such studies. At present, the MSRConf is the main conference in the area disseminating results of works on new processes, methods, techniques, tools, and ideas in this field [9]. Despite we know there are other venues about the topic (e.g. Empirical Software Engineering and Measurement, Empirical Software Engineering Journal), we focused on MSRConf that we believe is representative for this work.

In general, performing MSR studies is challenge for researchers since they need to deal with several software engineering artifacts, data sources, and understanding how better evaluate their purposes and results. Having a broad view of the area considering recent published studies may help new researchers to enter in the MSR community and also experienced researchers to perform works using new purposes, focuses, objects of analysis, or data sources [9].

In order to understand the MSR area and to identify its current targets, shortcomings, and gaps, we performed a systematic mapping study over studies published at five years of MSRConf (from 2010 to 2014). We chose this fresh period in order to investigate recent studies on MSR area. This allowed us to investigate how researchers are being conducted in this field and to address the following research question: "What are the current goals addressed by recent approaches in MSR?". Three complementing research questions were derived from the main question: (i) *RQ1*. Which are the main purposes, focus, and object of analysis of the identified approaches?; (ii) *RQ2*. Which are the data source types supported by mining software repositories and how are they evolving?; and (iii) *RQ3*. Which are the conducted empirical evaluations?

By answering these questions, we have identified that MSR approaches have been used for many different goals, mainly for comprehension of defects, analysis of contribution and behavior of developers, and software evolution comprehension. Some gaps were identified with respect to their goals, focus, and data source type (e.g. lack of usage of comments to identify smells, refactoring, and issues of software quality). Regarding the evaluation methods of the studies, our analysis pointed out to an extensive usage of some types of empirical evaluation. In order to make this research extensible and reproducible, we provide the extracted data in Section 3.

The remainder of this paper is structured as follows: Section 2 presents some related work. Next, Section 3 describes the systematic mapping method applied in this study. Section 4 discusses the main outcomes of the study and analyzes them. Section 5 considers the threats to the validity of this study. Finally, Section 6 presents a summary of the research and directions for future works.

2. RELATED WORK

Our systematic mapping study pointed out two relevant secondary studies that performed a general analysis of the studies in MSR field.

The first study, by Hammati et al. [9], reports best practices that the MSR community has developed over the past decade of research. It takes a different approach by creating a working cookbook (a set of best practices) that can be continuously used and updated as the MSR community matures and advances.

On the second study [10], the authors present a text mining exercise applied on the complete set of MSRConf papers to identify how the research on mining software repositories evolved in the last decade. The study describes an automatic and quantitative approach in order to address issues like trendy research topics, the most frequently (and less frequently) cited cases, and the most popular mining infrastructure [10].

These studies are different from the mapping study presented here. The first study extracted 268 comments from 117 full papers, categorizing them through grounded theory methodology. The second work performed an automatic text analysis, aiming to reflect the trend in MSR field. They provide a view of the area through different perspectives and do not follow a systematic process to select the papers and extract the data. Therefore, our

mapping study and the works discussed above are complementary to each other. Even that our study confirmed some trends pointed out by others, it also presents new findings.

3. SYSTEMATIC MAPPING METHOD

Some researchers have been working to provide stable methods for systematic literature review process [11][12][13][14]. In this study, we applied the approach defined by [12][15] in order to define the procedure of our mapping study.

We chose to conduct a mapping study because it allows analyzing the primary studies aiming to answer our research questions in a broader way and to collect evidence for directing future research. In addition, both data extraction and analysis are largely concerned with classification of the available studies. The steps of the performed mapping study are detailed in the following subsections.

3.1 Research Question

The main purpose of this work is to identify and characterize the set of recent primary studies published at the Working Conference on Mining Software Repositories (MSRConf) in order to identify the current picture of the MSR area. Thus, the main research question, which guides our study, is: "What are the current goals addressed by recent approaches in MSR area?".

To answer this question, we derived other three specific questions, described at the following:

RQ1. Which are the main purposes, focus, and object of analysis of the identified approaches?

This question aims to identify the goal of MSR approaches regarding facets based on classification proposed by Basili and Lambile [16]:

- **Purpose:** defines the main intention of study. The purpose of a study can be classified in characterizing, comprehension, evaluation, controlling, improving or prediction something [16]. Besides these categories, we also considered localization, classification, and identification, because they are frequently explored in the MSR field [17].
- **Focus:** is the main attribute of interest between purpose and object of analysis;
- **Object of analysis:** is the entity of interest in the analysis;
- **Point of view:** describes the perspective of a person needing the information;
- **Context:** describes the environment in which the analysis is performed.

Purpose, focus, and object of analysis are generally used together in MSR analysis. For example, to predict (purpose) defects (focus) analyzing bug reports (object of analysis) [3], and comprehending (purpose) power consumption (focus) analyzing mobile apps (object of analysis) [18].

In order to capture the goal of the studies, we focused on the first three facets for data extraction in our mapping study. Most papers did not explicitly mention what is the target audience for the approaches they propose (point of view) and the environment in which the analysis is performed (context).

¹ Studies based on MSR.

To perform the classification, we created a taxonomy based on proposed classifications [17][19] and grouped the different terms used by researchers for presenting purpose, focus, and object of analysis into categories. The classification scheme will be presented in Section IV, when we show the gathered data in the mapping study. The taxonomy used in our study is available at <http://goo.gl/hk67cc>.

RQ2. Which are the data source types supported by mining software repositories and how are they evolving?

This question aims to know which software mining infrastructures are used for MSR investigations and how they are evolving over time. According to [1], software mining infrastructure may be subdivided in three major categories: (i) historical repositories, such as source configuration management (SCM), bug tracking systems (BTS), and archived communications recording part of information about the evolution and progress of a project, (ii) runtime repositories, such as deployment logs, containing information about the execution and the usage of an application at a single or multiple deployment sites; and (iii) code repositories, such as Sourceforge.net, containing the source code of various applications developed by several developers.

Normally, source code repositories are used to access information about a specific software version (snapshots), or the software evolution history [10]. SCM systems are primarily used to register and maintain changes to source code artifacts. Moreover, SCM metadata has important information to support software comprehension. Examples of such information are: commit messages, commit date-time, commit author and changed data. BTS records maintenance changes and keeps track of the life cycle of bugs. This type of repository also may contain textual description of the bug, its location, and developers involved with the bug [4]. Archived communications store several types of discussions and requests for help among software team members. Examples of such repositories are mailing lists, questions and answers forums, and microblogging.

These repositories possess different information content, storage format, and vary in the way researchers mine their data [17]. To better analyze the data source type, we subdivided the repositories into two subcategories: structured and unstructured (Level 1). Level 2 (sublevel of Level 1) denotes the types of structured and unstructured repositories. It was subdivided in: (i) structured repositories: binary code, ITS, BTS, SCM metadata and source code, and (ii) unstructured repositories: mailing list, document, forum and log. In Level 3, we summarized the terms that characterize repository types of Level 2, such as "jar file" to represent "binary code" and "iBugs dataset" to represent "BTS".

RQ3. Which are the conducted empirical evaluations?

This question aims to identify whether the proposed approach has been evaluated through empirical methods, and if so, which method was used. We considered that a study has an empirical evaluation if it brings at least one section with some discussion dedicated to this topic. The empirical method is classified in one of the following types [21][22], depending on the purpose of the evaluation and how researchers described their evaluated methods:

Survey: is an empirical strategy to gather data from a population sample and to achieve an understanding of that population. The data collection process is achieved through questionnaires. The

facts, opinions, or beliefs of the people are polled to determine how the population reacts on a particular topic [21].

Case study: is a research methodology that studies a phenomena in its natural context [22]. We can find case studies that range from very ambitious and well-organized studies to small toy examples that claim to be case studies. In this sense, we respect the authors' assumption;

Controlled experiment: provides a formal, rigorous, and controlled investigation in which an intervention is introduced to observe its effects. Controlled experiments are normally used for identifying cause-effect relationships and the factors that may influence the outcome should also be controlled [23].

Feasibility Study: in this type of study, authors use the proposed approach over a software as a feasibility study and it only presents a proof of concept [20].

Exploratory Study: is conducted to analyze a problem that has not been defined. It focuses on exploring a specific context and does not intend to discuss final and conclusive solutions, but helps researchers to have a better understanding of the problem and generating ideas and hypotheses for new research [22].

Comparison with other tools: The authors performed a comparison based on checklist among their approach and another related tool [19].

3.2 Search strategy

We used a manual search to retrieve the studies for this mapping study. We considered full papers published at MSRConf from 2010 to 2014. We selected this conference because it is the main venue of this area. To ensure that results of the study selection are objective, we defined inclusion and exclusion criteria.

Inclusion criteria: The study need to be a full paper published at MSRConf on the selected period. Full papers present mature approaches, and likely more established studies. The study needs to explore a theory, a practice, or an approach related to the MSR process and software engineering domain.

Exclusion criteria: We excluded studies that do not address mining software repository and software engineering domain (e.g. [24]). Surveys and secondary experimental studies were removed, since they report approaches from others. Short papers, challenges, and showcases were also excluded (e.g. [9], [25]).

The inclusion and exclusion criteria over the five years of MSRConf resulted in 107 papers to be further analyzed and classified. The number of papers per year is as follows: 2010 – 15; 2011 – 19; 2012 – 17; 2013 and 2014 – 28 each.

3.3 Data Extraction Process

Two of the authors developed a taxonomy used in the data classification process. The first author performed a data extraction pilot with students from an experimental software engineering (ESE) discipline at a master and PhD course. In that moment, we discussed the application of the taxonomy in order to make clear the meaning of data.

After the pilot, two researchers have carefully read each paper and, in the next step, we have done a consensus process for each paper. If after that the researchers disagreed in a classification or extracted information, a third opinion was used to reach agreement.

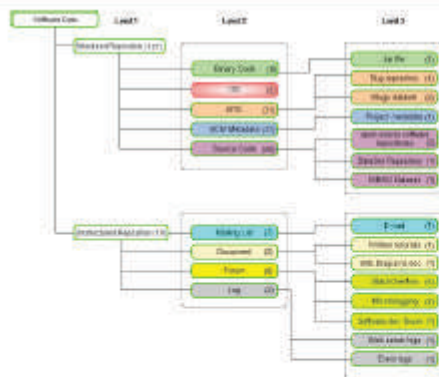


Figure 5. Data source types:

"BTS" with 31 studies - 21.5% are the most used software repositories. Similarly, structured repositories are more common with 127 studies. A possible reason may be that mining techniques are more challenging over unstructured data or that those repositories do not provide rich information to achieve MSR goals.

Other data sources we found are: (i) "iBugs dataset" with 2 studies, (ii) "Microblogging" with 2 studies, and (iii) "Stackoverflow" with 5 studies.

Figure 6 shows data sources distribution over the investigated years. We highlight two outcomes: (i) "source code", "SCM", and "BTS repositories" have been continuously used over the years; and (ii) the number of approaches using "archived communications" data source ("mailing list" and "forum") increased, reaching 12 studies in the last three years.

Finally, we identified several approaches combining more than one data source to achieve their goal. Figure 7 shows that 31.78%

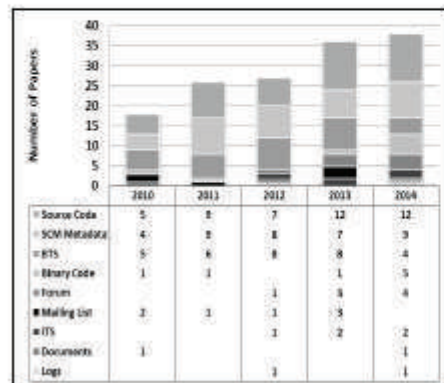


Figure 6. Data source by year

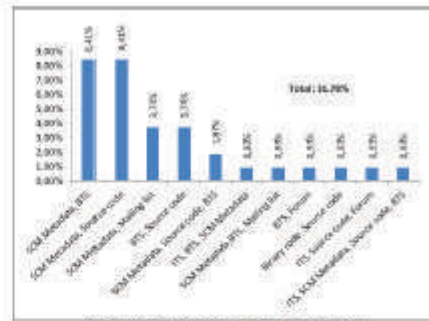


Figure 7. Studies per data source types

of the studies make joint use of different software repositories. The most recurrent combinations are "SCM metadata + BTS" and "SCM metadata + source code" representing 16.82% of all papers. An important note is that "SCM metadata" is present in both cases and in five more, totaling 63.64% of all combinations. Therefore, our analysis pointed out that "SCM metadata" performs an important role in MSR area. These results are aligned to results presented by Novais et al. [19].

RQ3: Which are the conducted empirical evaluations?

The studies analyzed in this mapping have conducted some types of empirical evaluation. Only one study did not conduct any evaluation [38].

We classified the studies according to six different evaluation methods. Figure 8 shows the distribution of each of them. More than half of the published studies had been evaluated through an "exploratory study". They represent about 58 studies - 51.8%. "Case study" represents 31 studies - 27.7% and "Controlled Experiment" with 8 studies - 7.1%. The rest of studies were evaluated through "comparison with other tools" and "feasibility study" with 5 studies - 4.5% each one, and "survey" with 4 studies - 3.6%.

Figure 9 summarizes evaluation methods performed per purpose. We can observe that exploratory study and case study are the main research methods used for evaluating "comprehension", "prediction", and "identification" tasks. An interesting point is

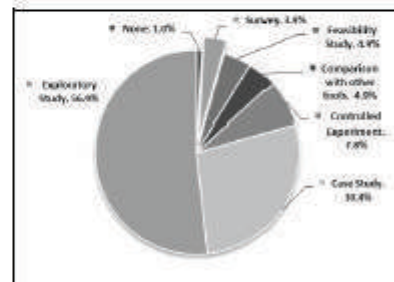


Figure 8. Evaluation methods:

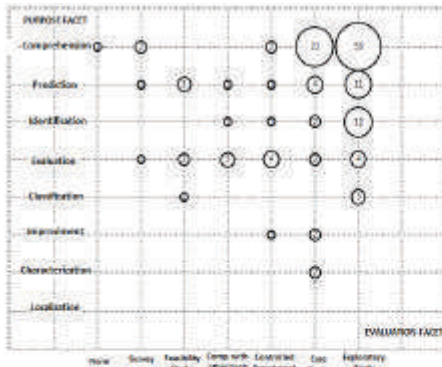


Figure 9. Purpose x evaluation methods

that "evaluation" and "prediction" were the only two categories where researchers have used all empirical methods to evaluate their approaches. Finally, we also can see that "survey" and "controlled experiments" have clearly been not much explored. In this sense, a deep analysis on those evaluation methods may be carried out, considering these types of studies.

5. THREATS TO VALIDITY

Selection Bias. At the start of the process, we applied the inclusion and exclusion criteria based on our judgment and the studies were included or excluded into mapping study. This means that some studies could have been categorized incorrectly. With the intention of mitigate this threat, we discussed the study protocol among the researchers to guarantee a common understanding. After that, two researchers read each paper and a third researcher was involved in the discussion when needed.

Construct validity. The research questions defined in this study may not completely cover the MSR area in spite of questions have addressed the main aspect of MSR field, as discussed by [18][20]. We defined facets based in Basili and Lamubile [16] in order to try answering the research questions.

Data extraction. Bias or problem of extraction can affect the classifications of facets and the analysis of selected studies. In order to reduce this bias, we discussed deeply the definitions of data items, the initial taxonomy, and our extraction form. After that, we performed a pilot data extraction and the doubts were discussed and solved in order to make clear the meaning of data items to other researchers. Finally, two researchers conducted in parallel and independently the data extraction process and mitigated occurred conflicts in a consensus meeting. If the researchers disagree about a classification or extracted information, we used a third opinion to resolve differences and reach agreement to make sure that the extracted data are valid and clear for further analysis.

External validity. We performed a systematic mapping study over studies published over five years of MSRConf. This point implies that we may have missed some relevant studies. Thus we cannot generalize our conclusions for whole MSRConf. However, our outcomes allow us to draw insights to guide further investigations.

6. CONCLUSION

In this paper, we performed a systematic mapping study on studies published at five years of MSRConf. We have extracted and analyzed data from 107 papers. Our findings show some trends on current use of purpose, focus, object of analysis, source data, and evaluate methods at MSRConf.

A large set of purposes and focus were used to investigate MSR artifacts, but "comprehension of defects" and "code" was the most used purpose, focus, and object of analysis in this field (RQ1). We defined a taxonomy that divided the data source types in structured and unstructured repositories. From this analysis, we might identify the main data source types in each category. Structured it more explored than unstructured repositories, but the number of approaches using unstructured data source is increasing in the last three years. Other software engineering artifacts are coming to be used, such as comments, emails, and microblogging. They have been analyzed alone or together with metrics extracted from structured repository to understand quality issues in software projects (RQ2).

Considering evaluation methods, we identified that almost all studies have performed some type of empirical evaluation and a low number of works has conducted "controlled experiments" and "surveys" (RQ3). Furthermore, other contribution of our study is that we make available the dataset of our data extraction and our classification (<http://goo.gl/hk67cc>).

As future work, we will investigate the gaps identified in this mapping study. One another research approach will be to conduct a broad evaluation of the proposed taxonomy by the researchers and software engineering practitioners. Finally, we also intend to investigate characteristics of software that may indicate quality problem in the development software, analyzing the code comments domain in order to design and evaluate new methods and tools to support Technical Debt identification.

7. ACKNOWLEDGEMENTS

This work was partially supported by CNPq Universal 2014 grant 458261/2014-9. The authors also would like to thank Mirella Sema for her support in the data extraction process.

8. REFERENCES

- [1] E. Hassan, "The road ahead for Mining Software Repositories," *Front. Softw. Maintenance, 2008. FoSM 2008.*, pp. 48–57, 2008.
- [2] G. Robles, U. Rey, and J. Carlos, "Replicating MSR: A study of the potential replicability of papers published in the Int. Work. Conf. Min. Softw. Repos.," pp. 171–180, 2010.
- [3] H. Naguib, N. Narayan, B. Briggs, and D. Helal, "Bug report assignee recommendation using activity profiles," *IEEE Int. Work. Conf. Min. Softw. Repos.*, pp. 22–30, 2013.
- [4] H. Valdivia Garcia, E. Shihab, and H. V. Garcia, "Characterizing and predicting blocking bugs in open source projects" *11th Work. Conf. Min. Softw. Repos.*, p.72–81, 2014.
- [5] F. Zhang, A. Mockus, I. Keivanloo, and Y. Zou, "Towards building a universal defect prediction model," *Proc. 11th Work. Conf. Min. Softw. Repos. - MSR*, pp. 182–191, 2014.

- [6] D. Steidl, B. Hammel, and E. Jürgens, "Incremental origin analysis of source code files," *Proc. 11th Work. Conf. Min. Softw. Repos. - MSR 2014*, pp. 42-51, 2014.
- [7] R. K. Saha, C. K. Roy, K. a. Schneider, and D. E. Perry, "Understanding the evolution of type-3 clones: An exploratory study," *IEEE Int. Work. Conf. Min. Softw. Repos.*, pp. 139-148, 2013.
- [8] J. Gil, M. Goldstein, and D. Moshkovich, "An empirical investigation of changes in some software properties over time," *Int. Work. Conf. Min. Softw. Repos.*, pp. 227-236, 2012.
- [9] H. Hemmati, S. Nadi, O. Beyral, O. Kononenko, W. Wang, R. Holmes, and M. W. Godfrey, "The MSR Cookbook: Mining a decade of research," 2013 10th Work. Conf. Min. Softw. Repos., pp. 343-352, May 2013.
- [10] S. Demeyer, A. Murgia, K. Wyckmans, and A. Lamkanfi, "Happy birthday! A trend analysis on past MSR papers," *Int. Work. Conf. Min. Softw. Repos.*, pp. 353-362, 2013.
- [11] B. Kitchenham, O. Pearl Brewster, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7-15, Jan. 2009.
- [12] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," pp. 1-10, 2007.
- [13] P. Brewster, B. a. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571-583, Apr. 2007.
- [14] C. Wohlin, P. Runeson, P. A. de Mota Silveira Neto, E. Engström, I. do Carmo Machado, and E. S. de Almeida, "On the reliability of mapping studies in software engineering," *J. Syst. Softw.*, vol. 86, no. 10, pp. 2594-2610, Oct. 2013.
- [15] S. E. Group, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.
- [16] V. R. Basili, F. Shull, and F. Lamblie, "Building knowledge through families of experiments," *IEEE Trans. Softw. Eng.*, vol. 25, no. 4, pp. 456-473, 1999.
- [17] H. Kagdi, M. L. Collard, and J. I. Maletic, "A survey and taxonomy of approaches for mining software repositories in the context of software evolution," pp. 77-131, 2007.
- [18] M. Linares-Vasquez, G. Bavota, C. Bernal-Cardenas, R. Oliveto, M. Di Penta, and D. Poshyvanyk, "Mining energy-greedy API usage patterns in Android apps: an empirical study," *Proc. 11th Work. Conf. Min. Softw. Repos. - MSR 2014*, pp. 2-11, 2014.
- [19] R. L. Novais, A. Torres, T. S. Mendes, M. Mendonça, and N. Zarwoka, "Software evolution visualization: A systematic mapping study," *Inf. Softw. Technol.*, vol. 55, no. 11, pp. 1860-1883, Nov. 2013.
- [20] A. Fernandez, E. Insfran, and S. Atrahalo, "Usability evaluation methods for the web: A systematic mapping study," *Inf. Softw. Technol.*, vol. 53, no. 8, pp. 789-817, 2011.
- [21] T. Punter, M. Ciolkowski, B. Freimut, and I. John, "Conducting on-line surveys in software engineering," 2003 Int. Symp. Empir. Softw. Eng. 2003. ISESE 2003. Proceedings., 2003.
- [22] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empir. Softw. Eng.*, vol. 14, no. 2, pp. 131-164, 2009.
- [23] D. I. K. Sjöberg, I. C. Society, J. E. Hannay, O. Hansen, and V. B. Kampenes, "A Survey of Controlled Experiments in Software Engineering," vol. 31, no. 9, pp. 733-753, 2005.
- [24] P. Rotella and S. Chulani, "Implementing quality metrics and goals at the corporate level," *Proc. - Int. Conf. Softw. Eng.*, pp. 113-122, 2011.
- [25] G. Robles, "Replicating MSR: A study of the potential replicability of papers published in the Mining Software Repositories Proceedings," *Proc. - Int. Conf. Softw. Eng.*, pp. 171-180, 2010.
- [26] R. Lotufo, L. Passos, and K. Czarniecki, "Towards improving bug tracking systems with game mechanisms," *IEEE Int. Work. Conf. Min. Softw. Repos.*, pp. 2-11, 2012.
- [27] X. Xia, D. Lo, X. Wang, and B. Zhou, "Tag recommendation in software information sites," *IEEE Int. Work. Conf. Min. Softw. Repos.*, pp. 287-296, 2013.
- [28] R. Stevens, J. Ganz, V. Filkov, P. Devanbu, and H. Chen, "Asking for (and about) permissions used by android apps," *IEEE Int. Work. Conf. Min. Softw. Repos.*, pp. 31-40, 2013.
- [29] G. Pinto, F. Castor, and Y. D. Liu, "Mining questions about software energy consumption," *Proc. 11th Work. Conf. Min. Softw. Repos. - MSR 2014*, pp. 22-31, 2014.
- [30] L. Ponzanelli, G. Bavota, M. Di Penta, R. Oliveto, and M. Lanza, "Mining StackOverflow to turn the IDE into a self-confident programming prompter," *Proc. 11th Work. Conf. Min. Softw. Repos. - MSR 2014*, pp. 102-111, 2014.
- [31] K. Bajaj, K. Panabiraman, and A. Meibah, "Mining questions asked by web developers," *Proc. 11th Work. Conf. Min. Softw. Repos. - MSR 2014*, pp. 112-121, 2014.
- [32] A. Nugroho, M. R. V. Chaudron, and E. Arisholm, "Assessing UML design metrics for predicting fault-prone classes in a Java system," *Proc. - Int. Conf. Softw. Eng.*, pp. 21-30, 2010.
- [33] S. Rao and A. Kak, "Retrieval from Software Libraries for Bug Localization: A Comparative Study of Generic and Composite Text Models," *Work*, pp. 43-52, 2011.
- [34] S. Karus and H. Gall, "A Study of Language Usage Evolution in Open Source Software," pp. 13-22, 2011.
- [35] J. Yang and L. Tan, "Inferring semantically related words from software context," 2012 9th IEEE Work. Conf. Min. Softw. Repos., pp. 161-170, Jun. 2012.
- [36] M. J. Howard, S. Gupta, L. Pollock, and K. Vijay-Shanker, "Automatically mining software-based, semantically-similar words from comment-code mappings," 2013 10th Work. Conf. Min. Softw. Repos., pp. 377-386, May 2013.
- [37] A. Murgia, P. Tourani, B. Adams, and M. Ortu, "Do developers feel emotions? an exploratory analysis of emotions in software artifacts," *Proc. 11th Work. Conf. Min. Softw. Repos. - MSR 2014*, pp. 262-271, 2014.
- [38] L. Nussbaum and S. Zaczarek, "The ultimate Debian database: Consolidating bazaar metadata for quality assurance and data mining," *Proc. - Int. Conf. Softw. Eng.*, pp. 52-61, 2010.