

UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

Daniela Soares Cruzes

**Análise Secundária de Estudos Experimentais em  
Engenharia de Software**

Campinas, São Paulo

2007

Daniela Soares Cruzes

**Análise Secundária de Estudos Experimentais em  
Engenharia de Software**

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Engenharia de Computação.

Orientador: Mario Jino

Co-orientadores: Manoel Gomes de Mendonça Neto  
Victor Robert Basili

Campinas, São Paulo

2007

Daniela Soares Cruzes

**Análise Secundária de Estudos Experimentais em  
Engenharia de Software**

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Engenharia de Computação.

Aprovação em 27/08/2007

Banca Examinadora:

Prof Dr. Mario Jino - UNICAMP

Prof. Dr. Eleri Cardozo - UNICAMP

Prof. Dr. Ivan Luiz Marques Ricarte - UNICAMP

Prof. Dr. Guilherme Horta Travassos - UFRJ

Profa. Dra. Sandra C. P. F. Fabbri - UFSCAR

Campinas, São Paulo

2007

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

C889a Cruzes, Daniela Soares  
Análise secundária de estudos experimentais em  
engenharia de software / Daniela Soares Cruzes. --  
Campinas, SP: [s.n.], 2007.

Orientadores: Mario Jino, Manoel Gomes de Mendonça  
Neto, Victor Robert Basili  
Tese (doutorado) - Universidade Estadual de Campinas,  
Faculdade de Engenharia Elétrica e de Computação.

1. Engenharia de software. 2. Mineração de dados  
(computação). I. Jino, Mario. II. Mendonça Neto, Manoel  
Gomes de. III. Basili, Victor Robert. IV. Universidade  
Estadual de Campinas. Faculdade de Engenharia Elétrica e  
de Computação. V. Título.

Título em Inglês: Secondary analysis of experimental software engineering

Palavras-chave em Inglês: Experimental software engineering, Vote counting,  
Data mining, Systematic reviews

Área de concentração: Engenharia de Computação

Titulação: Doutora em Engenharia Elétrica

Banca examinadora: Sandra Camargo Pinto Ferraz Fabbri, Guilherme Horta  
Travassos, Ivan Luiz Marques Ricarte, Eleri Cardozo

Data da defesa: 27/08/2007

Programa de Pós-Graduação: Engenharia Elétrica

## COMISSÃO JULGADORA - TESE DE DOUTORADO

**Candidata:** Daniela Soares Cruzes

**Data da Defesa:** 27 de agosto de 2007

**Título da Tese:** "Análise Secundária de Estudos Experimentais em Engenharia de Software"

Prof. Dr. Mario Jino (Presidente): \_\_\_\_\_

*Mario Jino*

Profa. Dra. Sandra Camargo Pinto Ferraz Fabbri: \_\_\_\_\_

*Sandra Camargo Pinto Ferraz Fabbri*

Prof. Dr. Guilherme Horta Travassos: \_\_\_\_\_

*Guilherme Horta Travassos*

Prof. Dr. Ivan Luiz Marques Ricarte: \_\_\_\_\_

*Ivan Luiz Marques Ricarte*

Prof. Dr. Eleri Cardozo: \_\_\_\_\_

*Eleri Cardozo*



## Resumo

Enquanto é claro que existem muitas fontes de variação de um contexto de desenvolvimento de software para outro, não é claro, a priori, quais variáveis específicas influenciarão a eficácia de um processo, de uma técnica ou de um método em um determinado contexto. Por esta razão, o conhecimento sobre a engenharia de software deve ser construído a partir de muitos estudos, executados tanto em contextos similares como em contextos diferentes entre si. Trabalhos precedentes discutiram como projetar estudos relacionados documentando tão precisamente quanto possível os valores de variáveis do contexto para assim poder compará-los com os valores observados em novos estudos. Esta abordagem é importante, porém argumentamos neste trabalho que uma abordagem oportunística também é prática.

A abordagem de análise secundária de estudos discutida neste trabalho (SecESE) visa combinar resultados de múltiplos estudos individuais realizados independentemente, permitindo a expansão do conhecimento experimental em engenharia de software. Usamos uma abordagem baseada na codificação da informação extraída dos artigos e dos dados experimentais em uma base estruturada. Esta base pode então ser minerada para extrair novos conhecimentos de maneira simples e flexível.

**Palavras-chave:** Engenharia de Software Experimental, Contagem de Votos, Mineração de Dados, Revisões Sistemáticas.

## Abstract

While it is clear that there are many sources of variation from one software development context to another, it is not clear a priori, what specific variables will influence the effectiveness of a process, technique, or method in a given context. For this reason, we argue that knowledge about software engineering must be built from many studies, in which related studies are run within similar contexts as well as very different ones. Previous works have discussed how to design related studies so as to document as precisely as possible the values of context variables and be able to compare with those observed in new studies. While such a planned approach is important, we argue that an opportunistic approach is also practical. This approach would combine results from multiple individual studies after the fact, enabling the expansion of empirical software engineering knowledge from large evidence bases.

In this dissertation, we describe a process to build empirical knowledge about software engineering. It uses an approach based on encoding the information extracted from papers and experimental data into a structured base. This base can then be mined to extract new knowledge from it in a simple and flexible way.

**Keywords:** Experimental Software Engineering, Vote Counting, Data Mining, Systematic Reviews.

## **Agradecimentos**

Todas as vezes que pensei em terminar esta tese, sempre achei que esta seria a página mais difícil de ser escrita. Eu vou começar os meus agradecimentos por Deus, no ano que conheci a Jesus como meu salvador, eu também comecei o doutorado, oito anos já se passaram. Tenho certeza de que Deus foi a presença certa em todos estes anos. Em todas as trocas de temas, de casa e de cidades. Obrigada Deus por estar presente sempre.

E por falar em presença, àqueles que sempre estiveram presentes mas ausentes em boa parte deste tempo. Os meus pais. Quanta saudade eu senti. Mas eles sempre me apoiaram, mesmo quando fui rebelde sem causa, eles sempre foram maravilhosos, sempre presentes na ausência. Nos momentos de angustia eu sabia que eles estavam lá, e que eu sempre teria o suporte de casa. Obrigada Papai e Mainha, vocês são o motivo da minha luta diária nestes anos todos. E pra vocês este título. Eu agora sou doutora!

Manuela, Gabi e Maria, nem sei o que dizer, vocês que souberam entender a minha ausência e os meus momentos de crise existencial. Obrigada por ainda estarem ai, e prontas sempre a me receber em casa novamente. Obrigada também ao carinho sempre de Mari, Cami e Jó, e do mais novo integrante da família, Fabiano.

As ausências foram sentidas mas as presenças também foram. Pessoas que entraram na minha vida como anjos, em cada momento eu precisava de um toque específico.

Tio Geovane, Tia Augusta, Aline e Marcos foram assim, entraram na minha vida no mestrado e me adotaram de uma maneira linda. Tem gente que pergunta se eu sou filha, às vezes digo que sou adotada, segundo Aline, meio irmã. É assim que eu me sinto. E reconheço quão importante vocês foram para a realização deste sonho. Este último ano então, ai meu Deus, foram vocês que aguentaram todos os momentos de stress que o ultimo ano trás. Obrigada por estarem aqui.

E o que dizer de Vicente e Meire. Acho que se eu estivesse em Salvador, teria sido adotada por eles. Quantos conselhos eu ouvi de Vicente. E quantas palavras de amizade e de

incentivo eu ouvi de Meire. Vocês conseguiram trazer alegrias nos dias de angustia. Obrigada por estarem aí.

Falando em adoção, o que dizer dos meus orientadores? Sou filha primeiramente do Jino, que nunca desistiu de mim, e que era a minha segurança mesmo quando parecia que tudo estava perdido. Manoel foi daqueles que entram por acaso, mas faz tanta diferença que muda o rumo das coisas, foi outro que me adotou. A bondade dele com a minha vida é tão grande que às vezes penso se mereço tanta bondade. E o Basili, sabe aquele sentimento de avô, me adotou como pai do Manoel, os tempos em Maryland foram os melhores da minha vida, o apoio sempre presente dele e as discussões foram extremamente importantes para nortear a tese. São três mentes brilhantes, que me deram a honra de conviver com eles durante algum tempo. Obrigada aos três, aprendi muito com vocês.

Os meus irmãos nesta árvore genealógica foram Forrest Shull e José Carlos Maldonado, os dois foram muito importantes em todo o processo. Todas as sugestões, idéias, comentários foram ouvidos e bem-vindos. Às vezes eu até fui chata com eles, procurando qualquer brecha que eles davam para que eu pudesse perguntar alguma coisa da tese. Obrigada pela paciência e o pronto atendimento sempre!

Eu sei que se eu for citar todos os nomes e descrever o quanto cada pessoa contribuiu para esta realização, eu escreveria mais duzentas páginas de agradecimentos, mas como não é possível, vou citar alguns nomes: Isaac, Atta, Paulinho, Chaim, Rosangela, Christiane, Jane, Tânia, Nanda, Raquel, André Godoy, Elias, Roberto Pinho, Rogério. Tenham certeza de que todos vocês tiveram uma participação especial.

Agradeço também aos professores que fazem parte da Banca Examinadora, que mesmo antes da defesa já contribuíram para este doutorado.

Agradeço ao Conselho Nacional de Pesquisa - CNPq - que me concedeu uma bolsa durante a realização deste doutorado, fato este que muito contribuiu para viabilização desta tese. À Unifacs, que me concedeu bolsa durante a estadia em Maryland. E à Universidade de Maryland e a Unicamp por terem sempre me recebido tão bem.

A Deus e a meus pais.

## Artigos Publicados

### Artigo publicado em periódico (Completo)

SHULL, Forrest ; CRUZES, Daniela Soares ; BASILI, Victor Robert ; MENDONÇA NETO, M. G. . Simulating Families of Studies to Build Confidence in Defect Hypotheses. *Information and Software Technology*, v. 47, n. 15, p. 1019-1032, 2005.

### Trabalhos em eventos (Completo)

CRUZES, Daniela Soares ; MENDONÇA NETO, M. G. ; MALDONADO, José Carlos ; JINO, Mario . Using Visualization to Bring Context Information to Software Engineering Model Building. In: *International Workshop on Visual Languages and Computing (VLC'2004)*, 2004, San Francisco. *Proceedings of the 2004 International Conference of Distributed Multimedia Systems*. Skokie, Illinois 60076 : Knowledge Systems Institute, 2004. v. 1. p. 219-224.

SANTANA, Christiane da Costa ; MENDONÇA NETO, M. G. ; CRUZES, Daniela Soares . Visualizing the Construction of Decision Trees Using *Treemaps*. In: *International Workshop on Visual Languages and Computing (VLC'2004)*, 2004, San Francisco. *Proceedings of the 2004 International Conference of Distributed Multimedia Systems*. Skokie, Illinois, 60076 : Knowledge Systems Institute, 2004. v. 1. p. 213-218.

MENDONÇA NETO, M. G. ; CRUZES, Daniela Soares ; SANTANA, Christiane da Costa . Interactive Construction of Classification Trees Using *Treemaps*. In: *CONFERENCIA LATINOAMERICANA DE INFORMÁTICA*, 2004, Arequipa, Peru. *ANALES DE LA XXX CONFERENCIA LATINOAMERICANA DE INFORMÁTICA*, 2004. v. 1. p. 1-10.

MENDONÇA NETO, M. G. ; CRUZES, Daniela Soares ; DIAS, Josimeire Machado ; OLIVEIRA, Maria Cristina Ferreira de . Using Observational Pilot Studies to Test and Improve Lab Packages. In: *International Symposium on Empirical Software Engineering*, 2006, Rio de Janeiro. *Proceedings of the 5th ACM-IEEE International Symposium on Empirical Software Engineering*, 2006. v. 1. p. 1-10.

CRUZES, Daniela Soares ; MENDONÇA NETO, M. G. ; BASILI, Victor Robert ; SHULL, Forrest ; JINO, Mario . Using Context Distance Measurement to Analyze Results across Studies. In: *International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2007, Madri, Espanha. *Proceedings of the 1st ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2007. v. 1. p. 1-10.

CRUZES, Daniela Soares ; MENDONÇA NETO, M. G. ; BASILI, Victor Robert ; SHULL, Forrest ; JINO, Mario. Extracting Information from Experimental Software Engineering Papers. *XXVI International Conference of the Chilean Computer Science Society, Event of the Chilean Computing Week*, November 8-9th - Iquique, Chile. v. 1. p. 1-10, 2007

## Resumos

CRUZES, Daniela Soares ; MENDONÇA NETO, M. G. ; SANTANA, Christiane da Costa . Aquisição de Conhecimento Durante a Mineração de Dados: Um Estudo de Caso em Tarefa de Classificação. In: Jornada Ibero-americana de Engenharia de Software e Engenharia de Conhecimento, 2004, Madri. Anales de las 4a Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería de Conocimiento. Madri, Espanha : Servicio de Publicaciones de la Facultad de Informática de la UPM ([www.fi.upm.es](http://www.fi.upm.es)), 2004. v. 2. p. 677-680

CRUZES, Daniela Soares ; MENDONÇA NETO, M. G. ; SANTANA, Christiane da Costa ; MALDONADO, José Carlos ; JINO, Mario . Construction of Predictive Software Engineering Models Using *Treemaps* Decisions Trees. In: Software Technology and Engineering Practice - STEP, 2004, Chicago. Site do Workshop on Predictive Software Models ([www.site.uottawa.ca/psm2004/](http://www.site.uottawa.ca/psm2004/)). Ottawa, Canada : School of Information Technology and Engineering, University of Ottawa, 2004. p. 1-4.

CRUZES, D. S. ; MENDONÇA NETO, Manoel Gomes de ; BASILI, V. ; SHULL, F. ; JINO, Mario . Automated Information Extraction from Empirical Software Engineering Literature: Is that possible?. In: International Symposium on Empirical Software Engineering and Measurement (ESEM), 2007, Madrid. Proceedings of the 1st ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2007. v. 1. p. 1-3.

# Índice

<b>CAPÍTULO 1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 MOTIVAÇÃO.....	1
1.2 VISÃO GERAL DA SOLUÇÃO .....	5
1.3 HISTÓRICO E AVALIAÇÃO DO TRABALHO .....	7
1.3.1 ETAPA 1: USO DE MINERAÇÃO DE DADOS PARA ANALISAR DADOS DE MÚLTIPLOS ESTUDOS. ....	7
1.3.2 ETAPA 2: ANÁLISE DE ESTUDOS - DADOS E ARTIGOS .....	9
1.3.3 ETAPA 3: FOCO NA EXTRAÇÃO DE INFORMAÇÕES DOS ARTIGOS.....	9
1.3.4 ETAPA 4: FOCO NA ANÁLISE DE CONTEXTOS EXPERIMENTAIS .....	10
1.3.5 ETAPA 5: ANÁLISE DO DESEMPENHO DE ESPECIALISTAS NA EXTRAÇÃO DE INFORMAÇÕES DOS ARTIGOS	10
1.3.6 ETAPA 6: AUTOMAÇÃO DA EXTRAÇÃO DE INFORMAÇÕES. ....	11
1.4 ORGANIZAÇÃO DESTE DOCUMENTO.....	11
<b>CAPÍTULO 2 REVISÃO DA LITERATURA E CONCEITOS BÁSICOS.....</b>	<b>12</b>
2.1 PROCESSO DE ENGENHARIA DE SOFTWARE EXPERIMENTAL .....	12
2.2 ANÁLISE PRIMÁRIA EM ENGENHARIA DE SOFTWARE EXPERIMENTAL .....	18
2.2.1 <i>Análise Estatística</i> .....	19
2.2.2 <i>Mineração de Dados em Engenharia de Software Experimental</i> .....	22
2.3 ANÁLISE SECUNDÁRIA EM ENGENHARIA DE SOFTWARE EXPERIMENTAL. ....	28
2.3.1 <i>Meta análise em ESE</i> .....	32
2.3.2 <i>Contagem de Votos (Vote Counting)</i> .....	37
2.3.3 <i>Revisão Sistemática em ES</i> .....	37
2.4 LIMITAÇÕES E VANTAGENS DOS MÉTODOS DE ANÁLISE SECUNDÁRIA .....	43
<b>CAPÍTULO 3 ABORDAGEM PARA ABSTRAÇÃO DE CONHECIMENTO A PARTIR DE ESTUDOS PRIMÁRIOS (SECESE) .....</b>	<b>47</b>
3.1 DEFINIÇÕES E VISÃO GERAL .....	47
3.2 A ABORDAGEM SECESE .....	50
3.3 PLANEJAMENTO DA ANÁLISE .....	53
3.3.1 <i>Definir objetivos de análise</i> .....	54
3.3.2 <i>Selecionar Artigos</i> .....	58
3.3.3 <i>Definir metodologia da análise</i> .....	61
3.4 AVALIAÇÃO DE QUALIDADE .....	62
3.5 EXTRAÇÃO DAS INFORMAÇÕES .....	65

3.5.1 Definições .....	65
3.5.2 Diretrizes Gerais para a Localização das Informações .....	66
3.5.3 Coleta de Resultados .....	67
3.5.4 Coleta de Informações de Contexto .....	69
3.6 ANÁLISE DE RESULTADOS.....	77
3.6.1 Formalização dos Resultados .....	78
3.6.2 Interpretação dos Resultados .....	82
3.7 ANOTAÇÕES.....	95
3.8 MANTENDO UMA BASE DE RESULTADOS .....	96
<b>CAPÍTULO 4 APOIO DE FERRAMENTAS.....</b>	<b>98</b>
4.1 INTRODUÇÃO .....	98
4.2 COLETA DE INFORMAÇÕES - FERRAMENTA INFOESE.....	99
4.2.1 Arquitetura e Modelo de Dados da Ferramenta InfoESE.....	99
4.2.2 Funcionalidades da Ferramenta InfoESE .....	101
4.3 ANÁLISE DOS RESULTADOS COM EXPLORAÇÃO VISUAL DE DADOS .....	114
4.3.1 A Técnica de Visualização para Exploração dos Resultados .....	114
4.3.2 A Ferramenta para Exploração dos Resultados.....	116
4.3.3 Exemplo de uso Ferramenta Treemap para Exploração da Base de Resultados. ....	117
4.4 APOIO DE FERRAMENTAS PARA ANÁLISE DE CONTEXTOS .....	121
4.4.1 Aglomeração Hierárquica de Dados .....	121
4.4.2 Aglomeração com a ferramenta HCE.....	122
4.4.3 Exemplo de uso Ferramenta HCE para Exploração da Base de Resultados .....	129
<b>CAPÍTULO 5 AVALIAÇÃO DA SECESE .....</b>	<b>136</b>
5.1 ESTUDO DE VIABILIDADE: ARTIGOS DE DETECÇÃO DE DEFEITOS .....	136
5.1.1 Detecção de Defeitos – Uma análise mais formal .....	139
5.1.2 Resultados da Análise.....	140
5.1.3 Uma Base de Resultados Expandida .....	144
5.1.4 Inclusão de artigos de outros autores.....	145
5.2 AVALIAÇÃO EXPERIMENTAL EM ARTIGOS GERAIS .....	149
5.2.1 Experimentadores .....	149
5.2.2 Participantes.....	150
5.2.3 Materiais.....	150
5.2.4 Procedimento.....	153

5.2.5 Coleta de Dados.....	154
5.2.6 Resultados/Lições Aprendidas.....	154
5.2.7 Ameaças à Validade.....	164
5.3 AVALIAÇÃO EXPERIMENTAL EM ARTIGOS DE TESTE.....	165
5.3.1 Experimentadores.....	165
5.3.2 Participantes.....	165
5.3.3 Materiais.....	167
5.3.4 Procedimento.....	168
5.3.5 Resultados/Lições Aprendidas.....	169
5.4 ENTREVISTA COM ESPECIALISTAS.....	175
5.5 EXPERIMENTO COM ESPECIALISTAS.....	178
5.5.1 Experimentadores.....	178
5.5.2 Participantes.....	178
5.5.3 Materiais.....	179
5.5.4 Procedimento.....	179
5.5.5 Resultados/Lições Aprendidas.....	180
5.6 VISÃO GERAL DOS ESTUDOS REALIZADOS.....	183
<b>CAPÍTULO 6 CONCLUSÃO.....</b>	<b>185</b>
6.1 PRINCIPAIS RESULTADOS.....	185
6.2 PRINCIPAIS CONTRIBUIÇÕES.....	188
6.3 LIMITAÇÕES DO TRABALHO.....	188
6.4 TRABALHOS FUTUROS.....	189
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>192</b>
<b>APÊNDICE A FORMULÁRIOS DOS EXPERIMENTOS.....</b>	<b>204</b>
<b>APÊNDICE B ENTREVISTA COM ESPECIALISTAS.....</b>	<b>218</b>
<i>B.1 - Entrevista com Especialista I.....</i>	<i>218</i>
<i>B.2 - Entrevista com Especialista II.....</i>	<i>223</i>
<b>APÊNDICE C NOTAÇÃO UTILIZADA NA MODELAGEM DA SECESE.....</b>	<b>229</b>
<i>C.1 – Notação.....</i>	<i>229</i>
<b>APÊNDICE D DICIONÁRIO DE DADOS DA BASE DE RESULTADOS.....</b>	<b>233</b>

## Lista de Figuras

FIGURA 1 - VISÃO GERAL DA SECESE .....	6
FIGURA 2 - CONCEITOS DE UM ESTUDO EXPERIMENTAL .....	14
FIGURA 3 - PROCESSO DE ENGENHARIA DE SOFTWARE EXPERIMENTAL .....	16
FIGURA 4 - PROCESSO PARA ANÁLISE ESTATÍSTICA DE DADOS EXPERIMENTAIS .....	20
FIGURA 5 - PROCESSO DE MINERAÇÃO DE DADOS .....	22
FIGURA 6 - MODELO DE REFERÊNCIA PARA VISUALIZAÇÃO [120].....	27
FIGURA 7 - ABORDAGEM DE KITCHENHAM PARA REVISÃO SISTEMÁTICA.....	39
FIGURA 8 – ABORDAGEM DE BIOLCHINI E OUTROS PARA REVISÃO SISTEMÁTICA.....	40
FIGURA 9 - DADOS, INFORMAÇÕES E CONHECIMENTOS COM RELAÇÃO À COMPREENSÃO E CONECTIVIDADE. ....	47
FIGURA 10 - EXEMPLOS DE DADOS, INFORMAÇÕES BRUTAS E RESULTADOS .....	49
FIGURA 11 - ABORDAGEM SECESE PARA ANÁLISE DE RESULTADOS EM ES.....	51
FIGURA 12 - DIAGRAMA DE FLUXOS DO PROCESSO DE SELEÇÃO DE ESTUDOS .....	61
FIGURA 13 - EXEMPLO DE AVALIAÇÃO DE QUALIDADE DOS ARTIGOS.....	64
FIGURA 14 - FORMULÁRIO DE DESCRIÇÃO DE CONTEXTO.....	70
FIGURA 15 - PROCESSO PARA INTERPRETAÇÃO DOS RESULTADOS .....	83
FIGURA 16 - ÁRVORE DE DISTÂNCIA BASEADA NAS INFORMAÇÕES DE CONTEXTO.....	91
FIGURA 17 – APOIO DE FERRAMENTAS PARA A SECESE.....	98
FIGURA 18 - MODELOS DE DADOS DA FERRAMENTA – INFOESE .....	100
FIGURA 19 – FERRAMENTA INFOESE: COLETA DE INFORMAÇÕES DE ANÁLISE .....	103
FIGURA 20 – FERRAMENTA INFOESE: COLETANDO OBJETIVOS DE ANÁLISE.....	104
FIGURA 21 – FERRAMENTA INFOESE: COLETA DE INFORMAÇÕES GERAIS SOBRE OS ARTIGOS .....	105
FIGURA 22 - FERRAMENTA INFOESE – COLETA DE INFORMAÇÕES GERAIS DO CONTEXTO DOS ESTUDOS .....	106
FIGURA 23 - FERRAMENTA INFOESE – COLETA DE VARIÁVEIS DEPENDENTES E INDEPENDENTES. ....	107
FIGURA 24 - FERRAMENTA INFOESE – COLETA DE INFORMAÇÕES SOBRE INSTRUMENTAÇÃO .....	108
FIGURA 25 - FERRAMENTA INFOESE – COLETA DE INFORMAÇÕES SOBRE OS PRODUTOS DE TRABALHO .....	109
FIGURA 26 - FERRAMENTA INFOESE – COLETA DE OUTRAS INFORMAÇÕES .....	110
FIGURA 27 - FERRAMENTA INFOESE – COLETA DE RESULTADOS.....	111
FIGURA 28 - FERRAMENTA INFOESE – FORMALIZAÇÃO DE RESULTADOS .....	112
FIGURA 29 – FERRAMENTA INFOESE – COLETA DE CONCLUSÕES DA ANÁLISE.....	113
FIGURA 30 – DESENHO TRADICIONAL DE UMA HIERARQUIA EM UMA ÁRVORE .....	115
FIGURA 31 – REPRESENTAÇÃO DO MAPA EM ÁRVORE.....	115
FIGURA 32 - FERRAMENTA TREEMAP.....	119
FIGURA 33 - FERRAMENTA TREEMAP- HIERARQUIA DE RESULTADOS .....	120

FIGURA 34 – AGLOMERAÇÃO HIERÁRQUICA E O DENDROGRAMA .....	122
FIGURA 35 – MOSAICO COLORIDO PARA UM CONJUNTO DE DADOS MULTIDIMENSIONAL. ....	124
FIGURA 36 – UM MOSAICO COLORIDO ANEXADO A UM DENDROGRAMA.....	125
FIGURA 37 – BARRA DE SIMILARIDADE MÍNIMA.....	126
FIGURA 38 – FERRAMENTA HCE – AGLOMERANDO CONTEXTOS .....	127
FIGURA 39 -- FERRAMENTA HCE –AGLOMERANDO CONTEXTOS BASEADOS NOS ATRIBUTOS TEÓRICOS .....	128
FIGURA 40 - TÉCNICAS ESTUDADAS EM CADA ARTIGO – EXEMPLO DE AGLOMERAÇÃO MANUAL.....	135
FIGURA 41 – CONJUNTO DE RESULTADOS – HIERARQUIA PRIMEIRO NÍVEL: VARIÁVEL DEPENDENTE E SEGUNDO NÍVEL VARIÁVEL INDEPENDENTE .....	142
FIGURA 42 – RESULTADOS QUE MOSTRAM A RELAÇÃO ENTRE DEFEITOS DE INTERFACE E DEFEITOS DE SOFTWARE.....	143
FIGURA 43 - RESULTADOS INDIVIDUAIS PARA O EXPERIMENTO I.....	156
FIGURA 44 - RESULTADOS INDIVIDUAIS EXPERIMENTO II.....	170

## Lista de Tabelas

TABELA 1 - ESTRATÉGIAS EXPERIMENTAIS .....	13
TABELA 2 – RESUMO DE ALGUNS MÉTODOS PARA SÍNTESE DE EVIDÊNCIAS QUALITATIVAS E QUANTITATIVAS.....	30
TABELA 3 - META ANÁLISE EM ES. ....	35
TABELA 4 - MACRO ATIVIDADES DA SECESE .....	52
TABELA 5 - ESTRUTURA PARA DETALHAR OBJETIVOS DE ANÁLISE.....	57
TABELA 6 – EXEMPLO DE DETALHAMENTO DE OBJETIVOS DE ANÁLISE.....	58
TABELA 7 – DOCUMENTAÇÃO DO PROCESSO DE BUSCA ADAPTADO DE [16].....	60
TABELA 8 - CRITÉRIOS DE QUALIDADE .....	63
TABELA 9 - RESUMO DA EXTRAÇÃO DE RESULTADOS DE ARTIGOS .....	68
TABELA 10 - ATIVIDADES DO PROCESSO PARA INTERPRETAÇÃO DOS RESULTADOS .....	83
TABELA 11 - OBJETIVOS DA VISUALIZAÇÃO X PROPÓSITOS DE ANÁLISE .....	86
TABELA 12 - POSSIBILIDADES DE CONCLUSÕES .....	93
TABELA 13 - ESTRUTURA PARA CADA TIPO DE ANOTAÇÃO. ....	96
TABELA 14- CRITÉRIOS DE QUALIDADE - FATOR COMO A PRÁTICA FOI APLICADA. ....	118
TABELA 15- CRITÉRIOS DE QUALIDADE - FATOR COMO A PRÁTICA FOI APLICADA. ....	132
TABELA 16 - GRUPOS DE ESTUDOS DEFINIDOS POR JURISTO ET. AL. ....	133
TABELA 17 – COMPARAÇÃO DOS RESULTADOS PRODUZIDOS ATRAVÉS DA SECESE E DA ABORDAGEM MANUAL POR MARICK.....	148
TABELA 18 - EXPERIÊNCIA DOS PARTICIPANTES - EXPERIMENTO I.....	150
TABELA 19 - PROJETO EXPERIMENTAL - EXPERIMENTO I .....	153
TABELA 20 - EXPECTATIVAS E RESULTADOS EXPERIMENTO 1 – COLETA DE RESULTADOS .....	156
TABELA 21 - EXPECTATIVAS E RESULTADOS - COLETA DE INFORMAÇÕES DE CONTEXTO - EXPERIMENTO 1 .....	158
TABELA 22- EXPERIÊNCIA DOS PARTICIPANTES DO EXPERIMENTO II.....	166
TABELA 23 - PROJETO EXPERIMENTAL DO EXPERIMENTO II.....	168
TABELA 24 - EXPECTATIVAS E RESULTADOS - COLETA DE INFORMAÇÕES DE CONTEXTO NO EXPERIMENTO II.....	172
TABELA 25 - EXPERIÊNCIA DOS PARTICIPANTES - EXPERIMENTO III.....	179
TABELA 26 - PROJETO EXPERIMENTAL DO EXPERIMENTO III.....	180
TABELA 27 - RESULTADOS DO EXPERIMENTO III - DIVIDIDO POR EXPERIÊNCIA .....	181
TABELA 28 - EXPECTATIVAS E RESULTADOS EXPERIMENTO III .....	181
TABELA 29 - EXPECTATIVAS E RESULTADOS - INFORMAÇÕES DE CONTEXTO - EXPERIMENTO III.....	182
TABELA 30 - ANÁLISE DE EFICIÊNCIA - EXPERIMENTO III.....	183



# Capítulo 1

## Introdução

### 1.1 Motivação

A Engenharia de Software (ES) é a área da computação que estuda as teorias, os métodos e as ferramentas usadas para desenvolver e manter programas. O aumento contínuo da complexidade e do tamanho dos produtos desenvolvidos faz com que o desenvolvimento e a manutenção de software se tornem custosos em termos de tempo e de esforço, e intensivo em termos de conhecimentos necessários para a sua execução. Muitos métodos e técnicas têm sido propostos para apoiar as atividades de ES, mas, não é raro encontrar métodos e técnicas que são avaliados inadequadamente e falham quando usados na prática.

A Engenharia de Software Experimental (ESE) é a disciplina da Engenharia de Software que investiga teorias, métodos e técnicas pela experimentação. As abordagens típicas para a avaliação experimental são estudos de caso, pesquisas de opinião e experimentos controlados [6, 29]. Um estudo experimental, em termos gerais, tem a finalidade de descobrir algo desconhecido ou de testar uma hipótese. Ele envolve um investigador que recolhe os dados e realiza uma análise para determinar o que os dados significam. Um experimento controlado é uma forma de estudo experimental na qual o investigador tem controle sobre os principais aspectos do estudo e as variáveis independentes que estão sendo estudadas. Ele pode também ser visto como uma operação realizada sob circunstâncias controladas a fim colocar à prova uma hipótese em observação [29].

Em um estudo experimental, os dados são coletados porque estão associados a uma teoria e têm um valor potencial. Uma teoria científica é um fato ou coleção de fatos estabelecidos ou experimentalmente verificados e relacionados ao domínio. Diferentemente do uso cotidiano da palavra teoria, esta não pode representar uma idéia não avaliada, ou apenas alguma especulação teórica. Sem prova, teorias representam apenas hipóteses. Assim, uma teoria é uma possível explicação para algum fenômeno, ou um padrão nos dados, que evidencia um relacionamento ou tendência em um atributo de interesse do pesquisador. Um cientista pode

---

analisar os dados para detectar esses relacionamentos ou tendências, denominados neste trabalho de resultados ou evidências. Tais padrões podem incluir a descoberta de que, por exemplo, à medida que o valor de uma propriedade (tal como o tamanho) aumenta, o de outra propriedade (tal como a confiabilidade) diminui. Um resultado é uma tentativa de explicação para determinados comportamentos, fenômenos, ou eventos que ocorrem nos dados. Um resultado define clara e concisamente o relacionamento esperado entre duas ou mais variáveis e define essas variáveis em termos operacionais e mensuráveis.

Estudos experimentais em Engenharia de Software têm sido conduzidos para investigar a análise, projeto, implementação, teste, manutenção, garantia de qualidade e reuso de software [19, 94]. Segundo Victor Basili, estamos vivendo a quarta fase da história da engenharia de software experimental [134]. No início (1974-1985), os pesquisadores executavam estudos isolados para propósitos específicos, usando, independentemente, estudos de caso e experimentos controlados como os meios para analisar uma questão particular de interesse. O foco principal nesta fase da ESE era a aprendizagem sobre o processo de medição e tentar identificar um conjunto apropriado de métricas para esta área [134].

Na segunda fase (1986-1999), iniciaram-se algumas tentativas de vincular os estudos. Os experimentos controlados, estudos de caso, quasi-experimentos e análises qualitativas tornaram-se parte de um cenário maior, cada um útil em sua própria vertente, mas voltados à finalidades diversas. Os experimentos controlados ganharam um valor especial na identificação de relacionamentos específicos entre variáveis específicas, enquanto os estudos de caso forneceram a oportunidade de aumentar a aplicabilidade dos resultados. Os pesquisadores aprenderam que poderiam reduzir o risco associado à experimentação executando estudos experimentais menores e usando a combinação dos estudos para construir a confiança em uma teoria baseada em tratamentos múltiplos. O foco principal era medir o relacionamento entre o processo e o produto. Entretanto, em Engenharia de Software os tipos de estudos executados e os tópicos estudados eram dependentes das oportunidades disponíveis [134].

Na terceira fase (2000-2004), os pesquisadores em ESE expandiram os estudos por diversos domínios e ambientes. Houve diversos exemplos de construção de conhecimento para um número limitado de técnicas em ambientes e domínios diferentes, isto é, estudando o efeito do contexto nas técnicas. Neste momento, os pesquisadores em ESE compreenderam melhor que o contexto pode mudar os resultados do efeito no estudo e que era difícil se identificar quais são as variáveis de contexto que têm o potencial de mudar estes resultados. Isto significa que a experimentação em engenharia de software deve envolver muitos investigadores, muitos ambientes, muitos domínios, e que não evoluiremos uma base de conhecimento sem colaboração. Isto implica que necessitamos definir focos porque coletar experiências em diversos ambientes, domínios, e tecnologias é uma tarefa muito difícil e custosa [134].

Para o presente e para o futuro, Víctor Basili diz que os focos de trabalho têm que ser limitados, restringindo o contexto da experimentação, construindo corpos de conhecimento sobre um domínio específico. Para cada domínio, isto envolve o recolhimento do folclore, as entrevistas, os estudos de caso, os experimentos controlados, as bases da experiência, etc. Segundo ele, a partir daí poderemos combinar o que foi aprendido para construir corpos de conhecimento maiores sobre diversos domínios [134].

Existe hoje, portanto, uma grande necessidade de se comparar resultados e derivar conclusões a partir de diferentes estudos experimentais. Há duas estratégias principais para construir essas conclusões. Uma é criar estudos experimentais relacionados, cuidadosamente planejados para ter seus resultados combinados. A segunda é analisar os resultados de estudos “não-relacionados”, isto é, estudos que não foram executados com o objetivo de ter seus resultados combinados. Esta segunda abordagem é denominada de estudo secundário, pois utiliza os resultados previamente obtidos em estudos primários.

Algumas estratégias usadas para execução de estudos primários relacionados são:

- Famílias de estudos [89, 132]: Os autores que advogam por esta abordagem acreditam que os dados de tais estudos podem ser combinados sem a perda de significância. As famílias de experimentos permitem que os pesquisadores respondam às perguntas que

---

estão além do escopo dos experimentos individuais. Assim, elas ajudam a generalizar resultados. Contudo, as famílias de estudos são caras em termos de tempo e esforço e, por este motivo, ainda é difícil gerar resultados gerais o bastante para criar um corpo de conhecimento. Além disto, alguns pesquisadores argumentam que problemas e limitações experimentais podem ser mais facilmente difundidos em famílias de experimentos relacionados [60]. Assim, as conclusões serão mais fracas do que se tivéssemos executado os estudos independentemente. Além disto, problemas no projeto de um experimento podem ser propagados para todos os outros estudos. E, mesmo nesta estratégia, o analista ainda precisa de uma abordagem para agregar resultados dos experimentos.

- Multi-Método [64]: Ao invés de usar os resultados obtidos com uso repetido do mesmo método de pesquisa, esta estratégia propõe que os investigadores triangulem os resultados obtidos a partir de métodos diferentes que não tenham fraquezas sobrepostas e que tenham pontos fortes complementares. Apesar de aplicações bem sucedidas [98], esta estratégia não é largamente usada em ES. Sua principal limitação está no alto custo de sua aplicação. O uso de multi-métodos tem um custo ainda maior do que o de realizar uma família de experimentos já que um universo mais heterogêneo de estudos tem que ser planejado e executado. Além disso, não se possui uma abordagem sistemática para se triangular os resultados de estudos heterogêneos.

A principal estratégia usada para execução de estudos secundários é agregação de estudos primários. Esta estratégia corresponde ao processo de organizar, resumir, e de generalizar resultados encontrados previamente - é também chamado de revisão integrativa ou sistemática.

- Agregação de Estudos: Uma revisão integrativa pode ser realizada de várias formas, das revisões narrativas à meta análise (quantitativa) [47, 51, 56], que surgiu como uma estratégia para inserir uma base estatística sólida às revisões narrativas da literatura [47]. Observamos nestas abordagens, especificamente naquelas que estão sendo adaptadas para a ES, que existe uma grande necessidade de se definir metodologias

estruturadas para se agregar ou generalizar eficientemente os resultados obtidos em estudos primários.

Esta tese propõe uma abordagem para combinar resultados de estudos primários diversos em engenharia de software, denominada SecESE (Análise Secundária em Engenharia de Software Experimental). A SecESE é inovadora, pois está focada especificamente na análise de resultados de ES, propondo uma abordagem genérica para analisar resultados experimentais, ao mesmo tempo que fornece diretrizes específicas para a caracterização e análise das variáveis de contexto típicas da área.

A abordagem SecESE tem duas fases principais. Na primeira fase extraem-se os resultados de diferentes estudos primários produzindo informação no mesmo nível de abstração a partir de todos os estudos envolvidos na análise. Na segunda fase analisam-se essas informações para extrair conclusões a partir de estudos experimentais diferentes (a análise secundária). A SecESE tem o potencial de construir um conjunto de conclusões que representam o corpo de conhecimento contido em múltiplos estudos primários, não necessariamente projetados para ter os seus resultados combinados. Além disso, tem o potencial de ser aplicado a qualquer tipo de estudo experimental em Engenharia de Software.

O objetivo principal deste trabalho é, portanto, propor e avaliar uma abordagem estruturada e métodos associados para agregar resultados de estudos experimentais em Engenharia de Software.

## **1.2 Visão Geral da Solução**

A abordagem de análise aqui proposta, SecESE, usa métodos explícitos e reproduzíveis que são executados sistematicamente (de acordo com um procedimento definido) e abertamente (assegurando que o procedimento da análise seja visível e reproduzível por outros pesquisadores). A Figura 1 mostra a abordagem proposta.

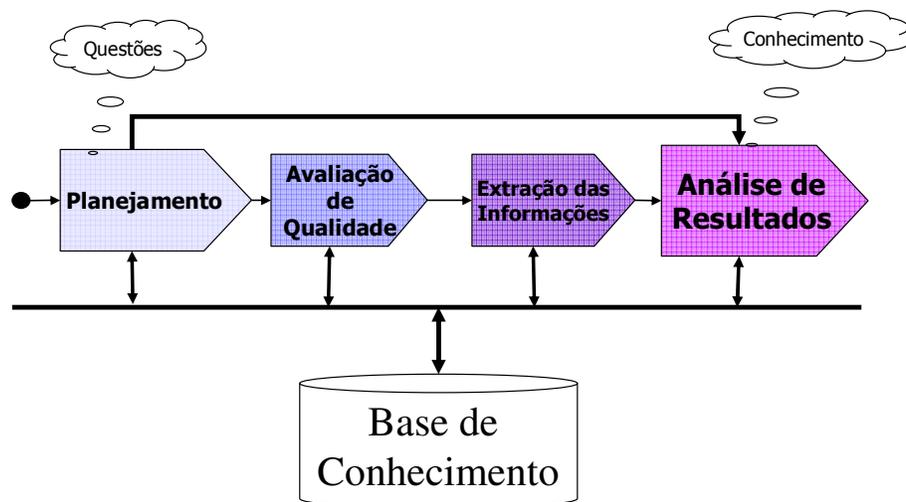
A abordagem tem uma atividade inicial de definição e planejamento da análise, na qual define-se uma estrutura padrão (template) para especificar os objetivos da análise, baseada no

---

método de Objetivos, Questões e Métricas (Goals, Questions and Metrics – GQM) [133]. Ainda neste planejamento existe a atividade de avaliação da qualidade dos objetos de análise. Esta atividade foi evidenciada no processo por ser uma atividade de grande importância para a garantia de que os resultados estão sendo devidamente avaliados.

A próxima atividade, extração de informações, pode ser realizada sobre dados ou sobre os relatórios dos estudos primários. O objetivo é criar um conjunto de resultados a partir das informações disponíveis nestes dois formatos.

Apesar do foco deste trabalho estar nas quatro atividades mostradas na Figura 1, não incluímos nesta tese o processo de extração de resultados a partir de dados de estudos primários em engenharia de software experimental. Parte do trabalho que fizemos nesta linha pode ser encontrada em artigos publicados na literatura [28, 34, 35, 36]. Para extração de resultados a partir de relatórios, definimos uma técnica específica. Esta técnica auxilia a criação de uma base de resultados, que é composta dos resultados propriamente ditos e das informações do contexto no qual esses resultados primários foram encontrados.



**Figura 1 - Visão Geral da SecESE**

Na última atividade da SecESE, define-se um processo que auxilia a análise dos resultados em diversos contextos experimentais de forma a se estabelecer conclusões sobre diversos

resultados. Uma das inovações da abordagem é o uso da mineração visual de dados para apoiar a análise dos resultados e dos relacionamentos entre os contextos em questão.

A saída do processo é um conjunto de conclusões que é adquirido pela execução do processo. A SecESE propõe também a criação de uma base estruturada de resultados e informações de contexto desses resultados. A vantagem de se criar tal base é a possibilidade de reuso e contínua expansão da base de resultados. Uma vez construída uma base de resultados, outros pesquisadores podem vir a modificá-la, expandi-la e reusá-la de acordo com seus objetivos de pesquisa.

### **1.3 Histórico e Avaliação do Trabalho**

Esta seção descreve a metodologia usada no desenvolvimento deste trabalho, descrevendo as seis etapas nas quais este trabalho foi desenvolvido.

#### **1.3.1 Etapa 1: Uso de mineração de dados para analisar dados de múltiplos estudos.**

A idéia de criar um processo para analisar dados de múltiplos estudos veio da necessidade de se analisar dados de um projeto denominado “Readers Project”. No escopo do projeto Readers, quatro replicações foram conduzidas para estudar a aplicação de técnicas de detecção de defeitos em documentos de requisitos de software [45, 46, 67, 68]. Analisou-se os dados de duas replicações usando um processo muito similar a um processo tradicional de mineração dos dados [127]. Algumas considerações sobre a análise feita nestas replicações são descritas em dois artigos [34, 36], nos quais iniciou-se a discussão sobre como usar as informações de contexto à análise de dados experimentais em Engenharia de Software e validação da qualidade de dados. Desenvolveu-se então uma ferramenta para auxiliar no processo de aplicação de mineração de dados em Engenharia de Software que implementou-se alguns conceitos necessários à construção e exploração visual interativa e iterativa de modelos em ES. Esta ferramenta foi o produto de uma dissertação de mestrado na Unifacs, orientada pelo professor Manoel Mendonça e co-orientada por mim [27].

---

Após este período sentiu-se a necessidade de se desenvolver experimentos para se entender melhor o processo de exploração visual de dados e melhor focar a sua utilização na análise de dados. Projetamos um experimento controlado para a avaliação de ferramentas de mineração visual de dados. Todo este processo está descrito em outro trabalho de mestrado na Unifacs, novamente orientado pelo professor Manoel Mendonça e co-orientada por mim [70]. Um artigo foi publicado no ISESE 2006 com alguns resultados desta experiência [86].

Surgiu então a necessidade de se aplicar um processo especializado de análise para dados experimentais em ES em outro domínio. O processo de análise de dados foi aplicado no contexto do projeto HPCS (“High Productivity Computing Systems”) na Universidade de Maryland – College Park. Este projeto envolve a colaboração entre pesquisadores experientes em estudos experimentais em Engenharia de Software e alguns pesquisadores da área de HPC (*High Productivity Computing*). O trabalho, concluído no final de 2005, concentrou-se nos estudos conduzidos para coletar dados. A equipe conduziu 12 estudos em salas de aula e 2 estudos observacionais no desenvolvimento de programas pequenos em HPC. Estes estudos focaram-se na medição de esforço, o fluxo de trabalho, e o tempo de execução dos programas. Eles são descritos mais detalhadamente na tese de doutorado de Lorin Hochstein [77]. Nosso foco neste projeto foi construir os modelos que descrevessem os efeitos das tecnologias, do tipo do problema, do tamanho do problema, e da experiência do programador, no desempenho e no esforço no desenvolvimento de programas. Este foi um ambiente adequado para melhorar partes do processo de análise de dados ligadas ao pré-processamento dos dados, a avaliação da sua qualidade e coleta estruturada de informação de contexto.

Neste estágio ficou muito clara a necessidade de se definir um método para combinar resultados de estudos diferentes. Havia-se definido um processo geral para analisar dados de estudos experimentais em ESE utilizando mineração de dados [28, 36], mas percebeu-se que um problema maior é a questão de como realizar “meta análise” em resultados de vários estudos experimentais; resolveu-se então focar nosso trabalho neste problema.

### **1.3.2 Etapa 2: Análise de estudos - dados e artigos**

Para avaliar as idéias da SecESE, fez-se um estudo para analisar resultados de alguns artigos, com o objetivo de combinar resultados e extrair o conhecimento neles contido [44]. Este estudo foi a primeira aplicação efetiva de todas as fases do processo proposto na Figura 1. Ele é apresentado como um estudo de viabilidade da abordagem nos capítulos subsequentes desta tese. A partir dessa avaliação, definiu-se dois aspectos importantes do processo de análise: um modelo estruturado para coleta de descrição do contexto de estudos experimentais apresentados em relatórios e o processo para analisar resultados oriundos de estudos experimentais diversos.

### **1.3.3 Etapa 3: Foco na extração de informações dos artigos**

A fase de extração de informações é uma das mais importantes e trabalhosas da SecESE. Esta fase precisava de uma investigação mais aprofundada, pois na primeira vez que o processo foi rodado, ele foi todo realizado por uma só pessoa. Não estava claro, portanto, a viabilidade de sua aplicação por outras pessoas.

Realizamos então dois experimentos para avaliar a extração das informações de artigos de maneira sistemática. O experimento foi executado em Maryland no âmbito de uma disciplina de pós-graduação em Engenharia de Software Experimental. Realizou-se o estudo em uma população de 7 estudantes, analisando artigos técnicos em diversas áreas de engenharia de software. Problemas foram detectados no material, na técnica e no projeto experimental. Corrigiu-se os problemas detectados no primeiro experimento, tornou-se a rodar o mesmo experimento controlado no Brasil, também em uma disciplina de pós-graduação em Engenharia de Software Experimental. Este estudo foi executado em uma população de 17 estudantes, utilizando artigos na área de testes de software que haviam sido previamente utilizados em um survey da área [100].

Após a realização do segundo experimento, fizemos uma entrevista semi-estruturada com dois dos três autores do survey previamente mencionado. Esta entrevista visou capturar como os autores extraíram as informações dos artigos e comparar esta abordagem manual com a

---

abordagem recomendada pela técnica de extração proposta neste trabalho. Os resultados indicam que a técnica de extração recomendada endereça os principais problemas enfrentados pelos autores na sua abordagem manual.

Todos os estudos e respectivos resultados estão descritos no Capítulo 5 desta tese.

#### **1.3.4 Etapa 4: Foco na análise de contextos experimentais**

A outra fase importante da SecESE é a análise das informações extraídas dos artigos. Quando realizou-se a etapa 2, focou-se mais na questão da análise de resultados. Porém, durante o desenvolvimento da etapa 3, elaborou-se um modelo para extração de informações dos artigos de forma padronizada. Com este modelo, pode-se desenvolver uma maneira de analisar os contextos extraídos dos artigos usando técnicas de agrupamento (clustering). Por meio deste modelo o analista consegue agrupar estudos de acordo com as similaridades e diferenças nos contextos. Pode-se também estabelecer pesos maiores nas características que o analista julga serem mais importantes na análise em questão. Realizou-se um estudo de viabilidade com alguns artigos da área de teste e o resultado foi positivo. Esta experiência é também discutida no Capítulo 5 e resultou na publicação de um artigo no ESEM 2007 [37].

#### **1.3.5 Etapa 5: Análise do desempenho de especialistas na extração de informações dos artigos**

Mesmo após a realização dos experimentos mencionados na etapa 2, julgou-se necessário avaliar a influência do desempenho de especialistas na extração das informações dos artigos. Estabeleceu-se então uma hipótese de que alunos inexperientes nem sempre conseguem extrair tudo o que existe num texto. Resolveu-se então realizar um experimento só com profissionais experientes para avaliar os a influência da experiência na eficiência e efetividade na extração de informações dos artigos. Este experimento é descrito no Capítulo 5.

### **1.3.6 Etapa 6: Automação da extração de informações.**

A etapa 6 está se iniciando no período final deste trabalho e seu desenvolvimento se dará em período posterior ao do doutoramento. Objetiva-se encontrar formas de automatizar a SecESE. Algumas atividades já possuem o apoio de ferramentas, outras não. Atualmente o foco dos esforços é na extração automática de informações dos textos. Atualmente, a área de ES não possui ferramentas específicas para extrair as informações que necessitamos. Entretanto, explorou-se com algum sucesso a utilização de ferramentas usadas para extração de informações em outros domínios, indicando a viabilidade desta abordagem para a ESE [38]. Dois trabalhos de mestrado estão se iniciando sob minha co-orientação na Unifacs, para propor soluções de mineração de textos que possibilitem a extração de informações diretamente de artigos de ESE.

## **1.4 Organização deste Documento**

Este documento está organizado da seguinte maneira. Este capítulo discutiu a motivação principal para o trabalho proposto além de descrever como ele foi desenvolvido. O Capítulo 2 apresenta os conceitos básicos e os termos que são usados no restante do trabalho. A Seção 2.1 apresenta os conceitos básicos do processo de ESE e as estratégias para execução de estudos experimentais. A Seção 2.2 contém um resumo das principais estratégias usadas na análise e interpretação de informações provenientes de estudos em ESE. O Capítulo 3 apresenta a SecESE, que é uma abordagem que combina a simplicidade e flexibilidade de análise de resultados usada em contagem de votos com a formalidade e aspectos de formalização que é usada em revisão sistemática. Um dos principais pontos de concentração da SecESE é a análise dos resultados. Definimos um método específico de contagem de votos apoiado por técnicas de visualização e aglomeração de informações. O Capítulo 4 detalha algumas ferramentas que podem ser usadas no apoio ao uso da SecESE. O Capítulo 5 descreve os estudos experimentais realizados para avaliar aspectos da SecESE. O Capítulo 6 contém as considerações finais e discute trabalhos futuros para prosseguimento da linha de pesquisa desta tese.

## **Capítulo 2**

### **Revisão da Literatura e Conceitos Básicos**

#### **2.1 Processo de Engenharia de Software Experimental**

Estudos experimentais tais como estudos de caso, experimentos ou pesquisas de opinião investigam as vantagens e desvantagens de métodos, técnicas e de ferramentas na Engenharia de Software. Estudos experimentais geram conhecimento experimental, que pode fazer parte do corpo geral de conhecimento da Engenharia de Software.

Tal conhecimento é necessário para ajudar organizações de Engenharia de Software a sobreviver em um mundo tão dinâmico quanto o atual. Estas organizações necessitam melhorar continuamente seus processos de desenvolvimento e de manutenção para ser competitivos. Elas também têm que melhorar a qualidade de seus produtos de software em um contexto de mudanças contínuas. Um outro elemento é o desenvolvimento tecnológico como, por exemplo, quando surgem novas tecnologias que criam oportunidades para novos mercados; e, ao mesmo tempo, impactam a forma de trabalho dos engenheiros de software, criando uma demanda por decisões de como utilizar tais tecnologias. Este é um contexto no qual os estudos experimentais devem ser realizados para avaliar como métodos e ferramentas podem ser aplicados na prática. Esses estudos também podem apoiar o crescimento da Engenharia de Software como uma disciplina madura de engenharia [19].

Existem diferentes abordagens de como instanciar um determinado estudo. Uma abordagem geral de realização de estudos experimentais é denominada estratégia da pesquisa ou estratégia experimental. As estratégias de pesquisa podem ser classificadas de muitas maneiras diferentes. Uma maneira amplamente usada que seguimos aqui é distinguir entre três estratégias principais: experimentos, estudos de caso e pesquisas de opinião [29]. Estas estratégias podem diferir, por exemplo, em projeto, procedimentos, coleta de dados, e métodos de análise. Uma definição destas estratégias, adaptada de [29], é dada na Tabela 1.

**Tabela 1 - Estratégias Experimentais**

<b>Estratégia</b>	<b>Definição</b>
Experimento	Em um experimento, uma investigação detalhada e formal é executada sob circunstâncias controladas. O objetivo é manipular uma ou mais variáveis e controlar as variáveis restantes. A finalidade de um experimento controlado é fazer observações cujas causas são não-ambíguas. Isto é alcançado isolando-se os efeitos de cada fator dos efeitos de outros fatores, podendo então estabelecer-se corretamente relações de causa e efeito [29].
Estudo de Caso	Em um estudo de caso, um projeto é monitorado e dados são coletados. Esta coleta de dados é realizada de acordo com um objetivo específico do projeto. Determinados atributos são monitorados (por exemplo, confiabilidade, custo) e dados são coletados com o propósito de medir estes atributos. Um estudo de caso é um estudo observacional enquanto um experimento é um estudo controlado. [94].
Pesquisa de Opinião	Uma pesquisa de opinião é uma investigação ampla onde a informação é coletada através de formulários padrão preenchido por um grupo de pessoas. Os meios principais de se recolher estes dados qualitativos ou quantitativos são entrevistas e questionários [29].

Existem outras classificações para as estratégias de pesquisa. Por exemplo, Zelkowitz e Wallace identificaram 12 modelos experimentais utilizados para avaliar novas tecnologias em computação: monitoração de projeto, estudos de caso, afirmação, estudos de campo, buscas na literatura, análise de dados legados, lições aprendidas, análise estática, repetição, experimento em ambiente sintético, análise dinâmica e simulação [94]. Kitchenham identificou nove tipos diferentes de estudos: experimento quantitativo, estudo de caso quantitativo, pesquisa de opinião quantitativa, seleção qualitativa, experimento qualitativo, estudo de caso qualitativo, pesquisa de opinião qualitativa, análise qualitativa de efeito, “benchmarking” [15]. Basili e outros distinguem os experimentos “in vivo” que são executados em uma organização e os experimentos “in vitro” que são conduzidas em um ambiente isolado e controlado de laboratório [135]. Além disso, um experimento pode ser executado com uma ou várias equipes, em um projeto ou em vários projetos. Travassos e Barros estenderam esta classificação com a inclusão de estudos “in virtuo” – aqueles conduzidos no laboratório/campo usando modelos computacionais do objeto de estudo e do ambiente, e “in-silico” – aqueles, conduzidos no laboratório com modelos computacionais

---

para participantes, objeto de estudo e ambiente, [54]. Existem ainda outras classificações [19].

Independente da estratégia de pesquisa, os elementos principais de um estudo experimental são as variáveis, os objetos, os participantes, o contexto do estudo, as hipóteses e o tipo de projeto do estudo [29, 53].

As variáveis de um estudo experimental podem ser dependentes ou independentes. As variáveis independentes referem-se à entrada do processo de experimentação. Essas variáveis podem também ser controladas e representam a causa que afeta o resultado do processo de experimentação. As variáveis dependentes referem-se à saída do processo de experimentação e representam o efeito provocado pelas variáveis independentes do estudo experimental. Como exemplo, suponha que queremos estudar o efeito de um método novo de desenvolvimento em relação à produtividade do pessoal. Podemos, por exemplo, ter escolhido introduzir um método de projeto orientado a objetos ao invés de um método funcional. A variável dependente no experimento é a produtividade. As variáveis independentes poderiam ser: o método do desenvolvimento, a experiência do pessoal, as ferramentas, e o ambiente de desenvolvimento. Para método de desenvolvimento como variável independente, pode-se usar dois tratamentos para esta variável: o método funcional e o orientado a objetos. A Figura 2, adaptada de [29], apresenta os relacionamentos entre os conceitos descritos acima.

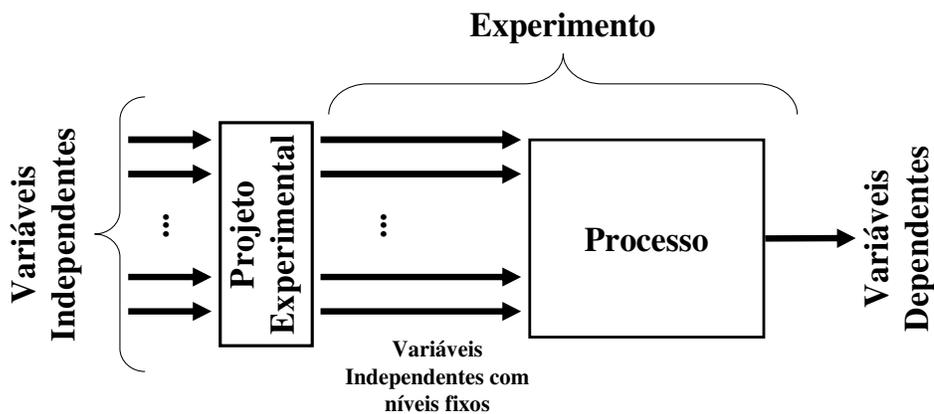


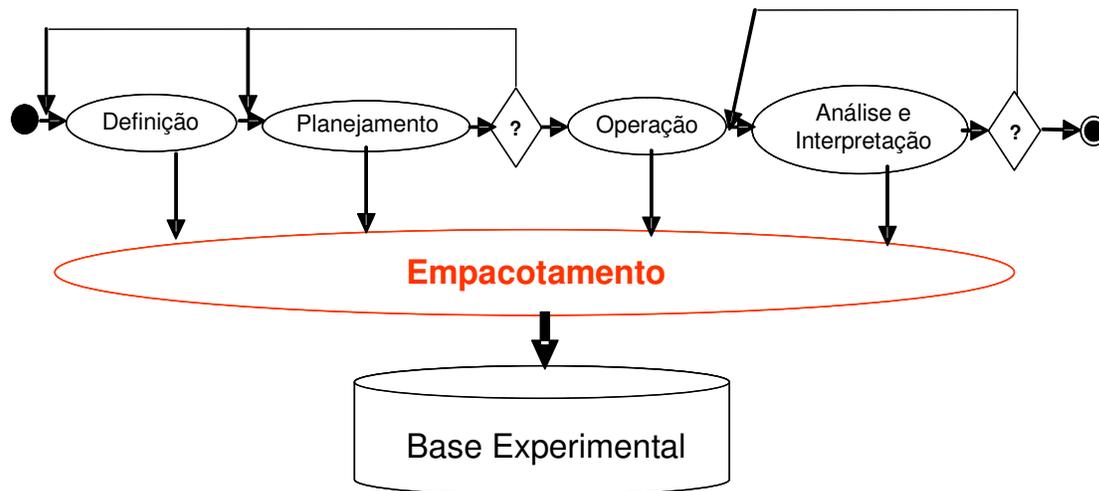
Figura 2 - Conceitos de um estudo experimental

Um conceito muito importante em estudos experimentais é o de contexto experimental. O contexto experimental é definido pelas condições de execução do estudo, pelos seus participantes, e seus instrumentos e ferramentas.

Um estudo experimental não é simples de ser realizado, é preciso preparar, conduzir e analisar o estudo apropriadamente para que o seu resultado seja válido. As atividades que geralmente devem ser executadas ao se conduzir um estudo experimental são descritas na Figura 3, adaptada de [53]. Estas atividades, discutidas em [19, 29, 53], são descritas a seguir:

- 1) *Definição do estudo*: O objetivo desta etapa é determinar o objetivo do estudo a ser executado. Baseando-se neste objetivo, uma estratégia experimental deve ser escolhida.
- 2) *Planejamento*: O objetivo desta etapa é fazer o delineamento do estudo experimental, produzindo, coletando e preparando todo o material que é requerido para conduzir o estudo experimental. Dependendo dos tipos de dados a serem coletados, o objetivo experimental pode ser expresso de uma maneira quantitativa, incluindo uma hipótese sobre o que esperar do estudo. Os dados quantitativos devem ser coletados e tratados a partir de questões que serão respondidas com entrevistas, questionários ou observações. Além disto, os métodos apropriados para a análise de dados devem ser selecionados. Esta seleção tem que considerar o tipo dos dados a serem tratados (por exemplo, quantitativo ou qualitativo) e o objetivo do estudo. Finalmente, o procedimento para conduzir a investigação experimental é planejado e apresentado em um plano de estudo. Geralmente, um teste piloto da execução é realizado para detectar e corrigir todas as deficiências nos materiais ou no projeto do estudo. Frequentemente é interessante que este “pré-teste” seja uma etapa à parte, com passos e critérios pré-estabelecidos para garantir a qualidade e validade do material experimental antes de se prosseguir para a fase de execução do experimento propriamente dito [86].

- 
- 3) *Execução*: O objetivo desta etapa é realizar o estudo de acordo com o planejamento e coletar os dados requeridos. Desvios do plano não devem ser tolerados e podem invalidar o estudo inteiro.
  - 4) *Análise*: O objetivo desta etapa é analisar os dados coletados para responder as questões do estudo. A análise é executada de acordo com os métodos de análise de dados selecionados durante a definição do estudo. Historicamente, esta fase emprega técnicas estatísticas para análise de dados.
  - 5) *Empacotamento*: O objetivo desta etapa é descrever o estudo, seus artefatos e seus resultados de modo que a comunidade externa seja capaz de compreender os resultados e seus contextos, bem como de replicar o estudo em um contexto diferente.



**Figura 3 - Processo de Engenharia de Software Experimental**

Basili e outros [132] afirmam que a experimentação em ES é difícil porque é complexa e consome tempo. Ela envolve fatores humanos, do domínio do problema, de processo e de recursos, tornando difícil mitigar as ameaças às validades internas e externas dos estudos experimentais [29, 41].

Esta dificuldade manifesta-se também na fase de análise. Uma razão importante para a experimentação em ES ser tão difícil é que os resultados de quase todo o processo dependem de um número potencialmente grande de variáveis dependentes do contexto, muitas delas não

controladas [19, 132]. Por causa disto, não podemos supor a priori que os resultados de um estudo se aplicam fora do ambiente específico em que ele foi executado. É difícil avaliar quão extensamente aplicáveis são os resultados de estudos isolados, mesmo quando estes estudos são bem executados [132]. Isto dificulta a compreensão da verdadeira contribuição destes resultados à área.

Alguns autores propõem o uso de famílias de experimentos como uma maneira de lidar com alguns desses problemas [132]. Porém, definir uma família de experimentos requer mais esforço do que preparar um estudo individual. Todavia, embora o esforço para preparar uma família de experimentos seja elevado os benefícios são grandes. Pesquisadores podem reusar o pacote experimental. Além da redução do esforço de projeto, o reuso pode ajudar também a melhorar a qualidade dos estudos. Ao mesmo tempo, os estudos individuais possuem um valor maior quando são parte de uma família de experimentos porque são analisados com respeito ao contexto da família inteira, não somente com respeito a seu próprio contexto. Isto é, as famílias de experimentos permitem um aprendizado mais abrangente que o de estudos individuais porque os resultados dos estudos são adicionados a um corpo de conhecimento, em vez de fornecer informações limitadas a um contexto isolado. Uma família de experimentos é mais do que uma série de experimentos replicados, porque a finalidade não é repetir a mesma experiência e avaliar seus resultados, mas examinar variações entre os experimentos [132]. Isto traz dificuldades para combinar seus resultados já que o contexto no qual os dados são coletados varia de um estudo para outro.

Outros autores sugerem o uso de uma abordagem “Multi-Método” na qual, em vez de usar os resultados obtidos com uso repetido do mesmo método de pesquisa, os pesquisadores devem triangular os resultados obtidos a partir de diferentes métodos, evitando sobreposição de fraquezas e complementando os pontos fortes dos estudos [64]. Embora a abordagem Multi-Método esteja fundamentada em um sólido argumento científico, a sua utilização prática ainda é bem limitada [98]. As razões para isto são o custo e a complexidade de se executar estudos com métodos diferentes de pesquisa, e a dificuldade para triangular seus resultados.

---

Para cada uma destas abordagens, existe uma grande necessidade de se definir metodologias estruturadas para agregar ou generalizar os resultados obtidos nos estudos.

## **2.2 Análise Primária em Engenharia de Software Experimental**

A análise primária é feita diretamente sobre os dados experimentais de um estudo individual. A análise de dados experimentais é realizada historicamente com a análise estatística [25, 29, 81]. Esta é a maneira mais madura de análise se o estudo envolve o teste de uma ou mais hipóteses iniciais.

Uma grande parte dos estudos experimentais em ES aplica testes estatísticos de hipóteses. Uma revisão de 72 estudos experimentais publicados de 1996 a 2003 nos volumes do “Journal of Empirical Software Engineering” mostrou que aproximadamente 65% destes estudos aplicaram testes estatísticos de algum tipo; por exemplo, teste de significância da diferença entre valores médios, de modelos alternativos ou teste de significância de uma variável incluída em um modelo de regressão [81].

Jørgensen e outros [81] afirmam que os contextos típicos usados em ES podem impedir o uso do teste estatístico de hipóteses e que o uso freqüente de testes estatísticos pode ter tido impactos indesejados em como os pesquisadores de ES projetam seus estudos. Eles discutem também que a aplicação do teste estatístico de hipóteses induz à formulação de questões de pesquisa e de projetos. Além disto, afirmam que a teoria de testes estatísticos de hipóteses é complexa e não há nenhuma interpretação unificada de todos os conceitos importantes. A Seção 2.2.1 descreve os usos de estatística em ESE, discutindo também as suas limitações.

A mineração de dados é apontada como uma opção para explorar dados experimentais em ES. Formalmente, a mineração dos dados pode ser definida como a extração de informação nova, útil e não-trivial de repositórios de dados [58, 88, 127]. A mineração dos dados representa uma mudança da análise baseada em verificação para a análise baseada em descobertas, da análise confirmatória para a análise exploratória.

O objetivo principal da análise de dados baseada em verificação é validar as hipóteses postuladas por analistas dos dados. Por outro lado, o objetivo da análise de dados baseada em

descobertas é automática ou semi-automaticamente, extrair informação nova e útil dos dados. Esta informação “útil” manifesta-se na forma de padrões, regras e modelos de predição, classificação e estimação. Há um grande número de técnicas que podem ser usadas para a análise de dados baseada em descobertas [22, 33, 48, 88, 90, 128].

Em engenharia de software grande parte dos trabalhos em análise de dados baseada em descobertas enfocam a construção de modelos. Muitos precedem até mesmo a era da mineração de dados. A maioria dos modelos de estimativas de custo pode ser classificada como baseados em descobertas porque foram modelos construídos diretamente dos dados. Mesmo após o crescimento da mineração de dados, em meados dos anos 90, a maioria dos artigos sobre análise de dados baseados em descoberta continuaram a se focar na construção de modelos para classificação, estimação e predição. Muitos artigos foram publicados sobre o uso de árvores da classificação, redes neurais e redes Bayesianas para a construção desses modelos [3, 17, 73, 76, 84, 96].

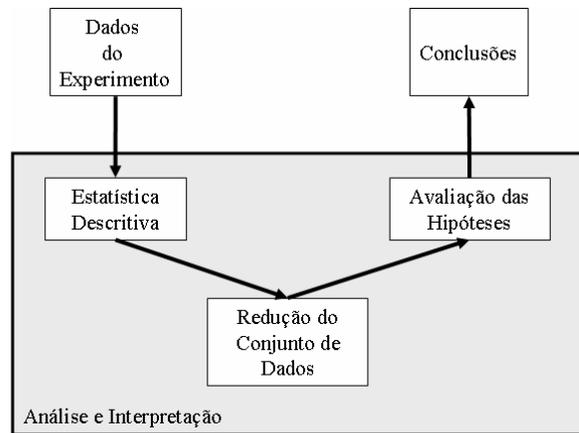
A análise de dados baseada em descobertas pode ser usada para extrair padrões, explorar o conhecimento de contexto, avaliar a qualidade dos dados e ajudar na combinação de resultados de estudos experimentais em ES. Acreditamos que esses conceitos têm que ser melhor explorados nos modelos para ES. Deve-se construir uma estrutura para o levantamento de dados e a análise que destaque a necessidade de coletar e considerar o conhecimento de contexto durante o processo de construção de modelos [34, 36]. A Seção 2.2.2 contém a descrição dos conceitos de mineração dos dados que utilizamos neste trabalho.

### **2.2.1 Análise Estatística**

A análise de dados experimentais é historicamente feita usando a análise estatística dos dados, normalmente utilizando técnicas de estatística inferencial [19, 29]. Esta é a maneira mais usual de análise, uma vez que o processo é dirigido por teste de uma ou mais hipóteses iniciais. Estas análises são executadas para a avaliação das hipóteses iniciais do estudo e geralmente consideram somente a informação disponível nos dados ou, às vezes, alguma

---

informação explícita que está codificada nos dados. O processo geralmente segue as etapas mostradas na Figura 4, adaptada de [29].



**Figura 4 - Processo para análise estatística de Dados Experimentais**

Na primeira etapa, os dados são caracterizados usando estatística descritiva, por meio de cálculos de centralidade, tendência e dispersão. Na segunda etapa, dados anômalos ou falsos são excluídos, reduzindo os dados a um conjunto de dados válidos. Na terceira etapa, os dados são analisados pelo teste das hipóteses e as hipóteses do estudo experimental são avaliadas estatisticamente, dado um nível de significância. Essas etapas são descritas mais detalhadamente por Wohlin e outros [29].

Estatística Descritiva é o nome dado ao conjunto de técnicas analíticas utilizadas para resumir o conjunto dos dados coletados em uma dada investigação a poucos números e gráficos. O objetivo é entender como a série de dados está distribuída, compreender melhor a natureza dos dados e identificar pontos de dados anômalos ou inválidos. Medidas de tendência central, tais como a mediana, média e a moda são usadas como indicadores que permitem que se tenha uma noção inicial de como os dados de um estudo experimental se comportam. Estas medidas não fornecem informações sobre a dispersão do conjunto de dados. Assim, é necessário medir o nível de variação da tendência central, isto é, analisar o quão dispersos ou concentrados são os dados (uma descrição mais completa dessas medidas é dada por Wohlin e outros [29]).

Na segunda etapa do processo mostrado na Figura 4, a redução do conjunto de dados consiste em identificar os dados anômalos baseado não somente na execução do estudo experimental, mas também inspecionando-se os resultados da execução nos formulários de dados coletados e considerando-se os resultados da estatística descritiva.

A última etapa, avaliação das hipóteses, é baseada em estatística inferencial. Ela se baseia no teste de hipóteses. O objetivo destes testes é verificar se é possível rejeitar uma determinada hipótese nula,  $H_0$ . Isto é, a hipótese nula ( $H_0$ ) descreve algumas propriedades da distribuição sobre a amostra e o experimentador quer rejeitá-la com certa significância. Para testar  $H_0$ , são definidos uma unidade de teste,  $t$ , e uma área crítica  $C$ . Isto significa que o teste de significância pode ser formulado da seguinte forma:

- If  $t \in C$ , rejeita-se  $H_0$ .
- If  $t \notin C$ , aceita-se  $H_0$ .

Os testes podem ser classificados em testes paramétricos e testes não-paramétricos [48]. Os testes paramétricos envolvem uma distribuição específica. Na maioria dos casos, pressupõe-se que alguns dos parâmetros envolvidos em um teste paramétrico estão distribuídos normalmente. Os testes paramétricos requerem também que os parâmetros possam ser medidos pelo menos em uma escala intervalar.[48]. Se os parâmetros não puderem ser medidos, isto significa, geralmente, que os testes paramétricos não podem ser usados. Nestes casos há um grande número de testes não-paramétricos que podem ser usados.

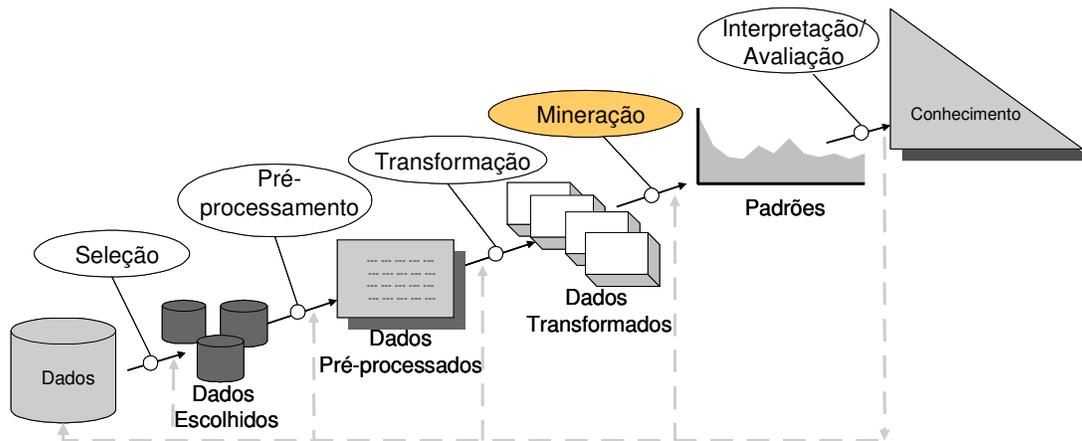
Os testes não-paramétricos não fazem o mesmo tipo de suposições dos testes paramétricos a respeito da distribuição dos parâmetros. Eles usam suposições mais gerais sobre as distribuições e por isto mesmo são válidos para um espectro maior de conjunto de dados. Por outro lado, os teste paramétricos tem níveis de significância mais difíceis de serem satisfeitos, exigindo amostras de dados maiores ou com menor dispersão.[48]

Seja qual for o tipo de estatística escolhida, paramétrica ou não-paramétrica, há testes especificamente destinados a amostras em que os fatores de variação são independentes e outros para amostras em que existe vinculação ou dependência entre os fatores. O

entendimento desses conceitos é importante para a análise dos resultados dos estudos experimentais dos artigos. Existem diversas referências para aplicação desses conceitos à engenharia de software experimental [29, 102].

## 2.2.2 Mineração de Dados em Engenharia de Software Experimental

A mineração dos dados foi definida como o processo de descoberta de informações previamente desconhecidas e potencialmente úteis em bases de dados [88, 127, 128]. Usamos aqui o conceito de que a mineração dos dados é o processo de induzir informação previamente desconhecida, potencialmente útil, de um conjunto de dados. A Figura 5 mostra um processo para mineração de dados, que consiste de cinco etapas [88, 127, 128].



**Figura 5 - Processo de Mineração de Dados**

A primeira etapa do processo de mineração consiste em selecionar os tipos de dados que serão usados pelo algoritmo de mineração. Conjuntos de dados brutos contêm geralmente uma variedade de dados que são desnecessários para se alcançar o objetivo da mineração de dados. Nesta etapa o analista dos dados tem que identificar onde os dados desejados estão e obter acesso a eles. Estas operações geralmente são conceitualmente simples mas de execução complexa. Muitas vezes, o acesso às fontes dos dados não é fácil. Além disto, os dados de ES geralmente não são organizados em bases de dados bem estruturadas e freqüentemente estão distribuídos sobre várias bases. Uma quantidade considerável de esforço é geralmente necessária para coletá-los e organizá-los em um lugar centralizado onde possam ser minerados.

A segunda etapa é o pré-processamento dos dados para a análise. Geralmente os dados têm que ser formatados, adaptados e, às vezes, transformados para a utilização por um algoritmo específico de mineração de dados. Formatar geralmente envolve o tratamento de dados ruins ou faltantes. A adaptação é necessária para preparar os dados para o algoritmo de mineração de dados. Alguns algoritmos, por exemplo, trabalham somente com dados nominais. Nestas situações, os dados numéricos têm que ser mapeados para dados categóricos baseado em algum critério consistente de adaptação.

Um pré-processamento mais sofisticado nos dados pode envolver a transformação dos dados. Esta operação é muito comum em dados de ES e ocorre quando o analista usa um conjunto de dados existentes para derivar novos atributos para a mineração dos dados. Por exemplo, suponha que uma organização de software tem uma base de dados de relatórios de erros. Entre estes dados estão: as datas de cada relatório e da remoção dos defeitos em um conjunto de projetos do software. Uma possibilidade de transformação dos dados seria usar as datas para calcular o atributo “severidade do defeito” em função do tempo gasto entre o momento de descoberta e o momento de reparo do mesmo.

Após o pré-processamento, os dados estão finalmente prontos para serem minerados por um algoritmo. A etapa de mineração de dados tem como objetivo extrair desses dados informações potencialmente úteis, desconhecidas e não-triviais. Nesta fase, é importante a utilização dos conhecimentos obtidos durante as fases anteriores em relação ao domínio dos dados e ao processo de coleta dos mesmos, de forma a melhor direcionar o algoritmo de mineração. Esta etapa pode envolver técnicas de mineração de dados muito diferentes. Pode-se, por exemplo, usar os dados pré-processados para produzir automaticamente gráficos contendo associações interessantes entre as características do projeto e os perfis dos defeitos do software; ou, obter uma visão geral das características e dos perfis dos defeitos usando ferramentas interativas de visualização de dados.

De acordo com Mendonça e Sunderhaft [88], as tarefas de mineração de dados são tipos básicos de operações que podem ser realizadas para a descoberta de diferentes tipos de conhecimento. Segundo Fayyad e outros [128], as principais metas das tarefas de mineração

---

são: a previsão, que utiliza variáveis para prever valores desconhecidos ou futuros de outras variáveis de interesse; e a descrição, que envolve a descoberta de padrões que descrevam os dados. As tarefas de mineração de dados são classificadas em seis categorias principais [88]:

- 1) Estimação. Consiste em examinar atributos de um conjunto de entidades (produtos, processos, e recursos) e, baseando-se em seus valores, atribuir valores a outro atributo desconhecido. Um exemplo típico de uma tarefa de estimação é usar os atributos que caracterizam um projeto para estimar seus custos.
- 2) Classificação. A classificação consiste em examinar os atributos de uma certa entidade e, baseando-se nos valores de cada atributo, atribuir-lhe uma categoria ou classe pré-definida.
- 3) Descobertas de Aglomerados (ou clustering). Visa à segmentação de uma população heterogênea em subgrupos (ou clusters) homogêneos. A aglomeração difere da classificação porque não se baseia em classes de dados pré-definidas. Os aglomerados são atribuídos pelo algoritmo de mineração.
- 4) Descobertas de associação. Consistem em identificar, em um conjunto de dados, os atributos que têm uma relação, ou são mais frequentemente associados, entre si. Um exemplo típico seria tentar encontrar os atributos da equipe de desenvolvimento do software (experiência, treinamento, conhecimento do domínio, etc.) que estejam associados aos atributos finais do produto de software (usabilidade, confiabilidade, manutenibilidade, etc.).
- 5) Visualização dos dados. É a tarefa de descrever informações complexas por meio de cenas visuais facilmente interpretáveis para um perito de um certo domínio. A visualização é baseada na premissa de que uma boa descrição de uma entidade (seja um recurso, um produto, ou um processo do software) facilitará a compreensão desta entidade e de seu comportamento.
- 6) Exploração interativa dos dados. Consiste em inspecionar grandes volumes de dados por meio de consultas usando controles interativos. Este tipo de tarefa é

freqüentemente combinada com visualização de dados, usando controles gráficos que permitem a formulação de “consultas” com movimentos do mouse. Esta abordagem, chamada de exploração visual de dados, permite que peritos em um certo domínio de conhecimento possam rapidamente explorar por si próprios grandes conjuntos de dados.

A última etapa do processo de mineração de dados, mostrado na Figura 5, é a assimilação da informação minerada. Isto é feito pela interpretação e assimilação da informação que a técnica de mineração de dados considera “interessante”. Funções de interessantismo são normalmente usadas para determinar o quanto um padrão pode de ser interessante para um perito [42]. Um padrão é interessante se ele é: improvável, novo, valioso e interpretável. No caso de construção de modelos, esta etapa consiste em avaliar a robustez e a eficácia dos modelos produzidos usando técnicas tais como a validação cruzada. Se aprovado, o modelo pode ser incorporado à base de conhecimento.

O processo de mineração de dados é, geralmente, altamente interativo. O analista dos dados tem que repensar a seleção dos dados sempre que a informação minerada não é interessante. Além disto, ele tem que refinar o pré-processamento dos dados se estes não estiverem completamente adequados ao algoritmo de mineração. O algoritmo necessita freqüentemente ser reajustado quando poucos fatos interessantes são identificados na etapa de assimilação dos dados. Por sua natureza, o processo de mineração de dados tem várias iterações. A qualidade da informação descoberta depende da qualidade dos dados brutos, do algoritmo de mineração, e de outras técnicas que são usadas durante todo o processo. Ele é também muito dependente da maneira pela qual o processo é conduzido e do quão bem os analistas dos dados interagem nas suas etapas.

#### 2.2.2.1 Exploração Visual de Dados

A visualização de informações é definida por Card [120] como a utilização de apoio computacional na representação visual de dados a fim de ampliar o conhecimento humano. Segundo ele as representações gráficas têm dois objetivos básicos: a comunicação de uma

---

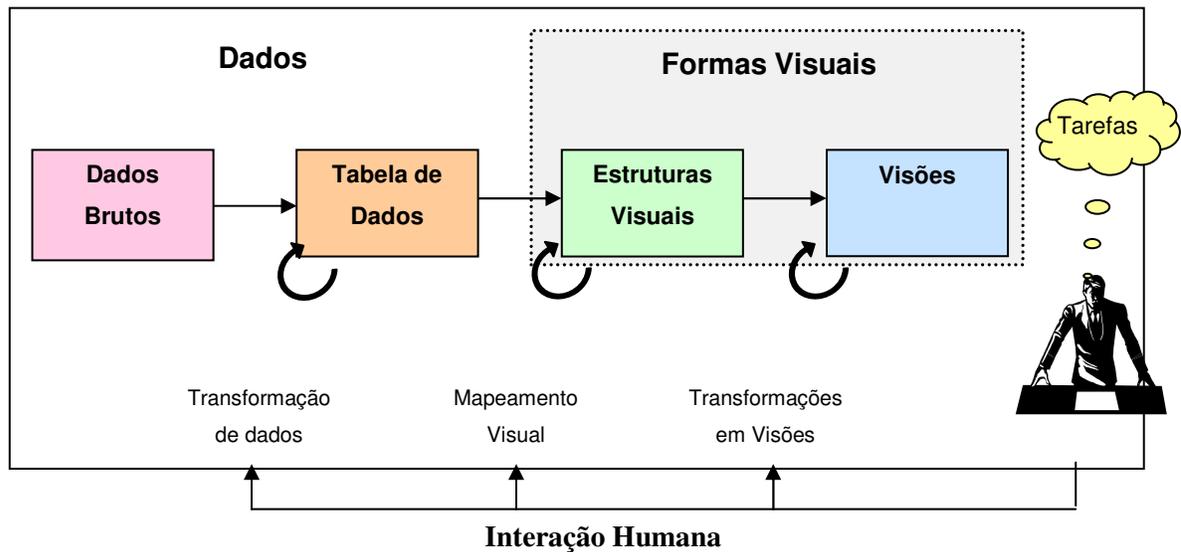
idéia pré-concebida e a utilização gráfica para criação ou descoberta de idéias. Apesar de as apresentações gráficas já serem utilizadas há bastante tempo, a visualização de informações tem se beneficiado da constante evolução dos computadores, que torna possível a manipulação de bases de dados cada vez maiores em tempo real e com baixos custos. Isto tem permitido a criação de novos métodos de visualização para o auxílio ao desenvolvimento cognitivo humano.

Uma técnica de visualização de informações é baseada numa representação visual e em mecanismos de interação que possibilitam ao usuário manipular a representação de forma a melhor compreender o que está sendo apresentado [23]. Segundo Keim [33] o uso de técnicas de visualização de informações tem basicamente três objetivos:

- Análise Exploratória – Inicia-se sem nenhuma hipótese sobre os dados e, de forma interativa, busca-se por estruturas e tendências. Esse processo resulta em hipóteses sobre os dados;
- Análise Confirmatória – Inicia-se com hipóteses sobre os dados e direciona-se a análise para testar estas hipóteses. O resultado pode ser a confirmação ou rejeição das hipóteses;
- Apresentação – Fatos conhecidos são escolhidos para serem apresentados, escolhe-se uma técnica apropriada para a apresentação e tem-se com isso uma apresentação eficiente destes dados.

Um modelo de referência do processo de visualização de informações pode auxiliar na escolha da técnica para visualização ou no desenvolvimento de novas técnicas. Card [120] afirma que a visualização é um mapeamento dos dados para uma forma visual que o homem possa interpretar. A Figura 6, adaptada de Card, descreve os mapeamentos usados para transformar dados em cenas visuais e serve como modelo de referência para visualização.

Neste modelo de referência, as setas que saem da esquerda para a direita indicam uma série de transformações nos dados. Cada seta pode indicar múltiplas transformações. As setas que saem da direita para a esquerda e as recursivas indicam o ajuste destas transformações por meio de controles operados pelo usuário.



**Figura 6 - Modelo de Referência para Visualização [120]**

Na “Transformação dos dados”, os dados brutos que podem se encontrar nas mais variadas formas, como planilhas ou textos, são mapeados em “Tabelas de dados” contendo apenas os dados de interesse do usuário. No contexto de engenharia de software experimental, estas tabelas de dados são geralmente descrições relacionais dos dados experimentais.

O “Mapeamento visual” transforma as “Tabelas de dados” em “Estruturas visuais”, mapeando valores nas tabelas de dados (atributos reais) para atributos visuais representáveis na tela bidimensional de um computador. O mapeamento visual é feito sobre um ou mais técnicas de visualização, tais como um histograma ou um diagrama de dispersão, que usa atributos visuais tais como os eixos X e Y do plano cartesiano ou marcas, cores e outras propriedades gráficas, para apresentar visualmente os vários campos (atributos reais) da tabela de dados.

Finalmente, as “Transformações em visões” são usadas para apresentar a cena visual a partir da especificação dinâmica de parâmetros gráficos como perspectiva, aproximação (zoom) e posicionamento do mapeamento visual.

O núcleo deste modelo de referência é o mapeamento da tabela de dados para uma estrutura visual, a qual é baseada em uma técnica de visualização com propriedades gráficas

---

adequadas ao domínio dos dados e de fácil entendimento pelo ser humano que vai analisá-los. Vários autores, na tentativa de criar melhores estruturas visuais e visões mais adequadas a tipos de dados específicos, desenvolveram diferentes técnicas de visualização de informações. Existem várias classes de técnicas de visualização de informações, entre elas podemos citar as técnicas: Orientadas a Pixel [33, 97]; Iconográficas [55, 110]; Hierárquicas [18, 20, 50] e Baseadas em Grafos [85, 104].

Para a análise primária de estudos experimentais em engenharia de software, propomos um processo específico para a análise baseada em exploração visual e mineração de dados [34, 35, 36]. Argumentamos nestes artigos que a construção de modelos é a ponta do iceberg, isto é, onde queremos chegar com a coleta e análise de dados. Todavia, para chegar lá, assuntos como qualidade de dados e conhecimento de contexto devem ser explorados. Nossos artigos discutem como a mineração de dados, a extração de padrões, visualização de informação e a exploração interativa de dados podem ajudar nesta tarefa. Também apresentamos um arcabouço para coleta e análise de dados que realça a necessidade de se coletar e considerar conhecimento de contexto durante o processo de construção do modelo.

No trabalho desenvolvido nesta tese, propomos a análise secundária por meio do uso de aglomerações e de mineração visual dos dados para analisar os resultados obtidos em estudos primários na engenharia de software experimental.

### **2.3 Análise Secundária em Engenharia de Software Experimental.**

O paradigma experimental em ciência é baseado na observação e na experimentação sistemática sobre os elementos que interagem no universo em questão. Como em vários campos da ciência baseados em experimentação, a Engenharia de Software Experimental contém dois tipos de investigação: estudos primários e secundários [66].

Os estudos primários são projetos específicos de processos metodológicos, dirigidos para avaliar diretamente a hipótese formulada pelo pesquisador, testando-a sob condições estabelecidas no controle metodológico observacional ou experimental. Os estudos secundários são aqueles produzidos para estabelecer comparações, generalizações

sistemáticas ou montagem entre as investigações individuais, selecionadas cientificamente de um conjunto de estudos primários [66].

A análise secundária é portanto a análise que é realizada nos resultados da análise primária dos estudos experimentais com o objetivo de estabelecer comparações, generalizações, sumarizações entre estudos primários individuais.

O interesse por estudos secundários na Engenharia de Software tem crescido significativamente [14, 32, 82], e este crescimento deveu-se à incorporação de técnicas desenvolvidas em outros campos, como a medicina. Existem diversos métodos para síntese de estudos experimentais. Estes métodos são sumarizadas brevemente na Tabela 2, adaptada de [95]. Entre estes métodos, a meta análise, a contagem de votos e as revisões sistemáticas têm sido adaptados para a Engenharia de Software nos últimos anos.

Para combinar resultados de vários estudos, a meta análise oferece um conjunto de técnicas quantitativas que permitem a sintetização dos resultados de muitos tipos de pesquisa, incluindo pesquisas de opinião, experimentos e quasi-experimentos [2, 7, 59, 75, 109, 118].

Uma meta análise utiliza métodos estatísticos para combinar os resultados de um número de estudos experimentais que examinaram a mesma questão de pesquisa, numa tentativa de sumarizar a totalidade de evidências ou resultados relacionados a uma questão em particular. Na meta análise o pesquisador define pelo menos um indicador do relacionamento sob investigação a partir de cada um dos estudos. Os indicadores são usados para calcular médias, desvios padrão e outras medidas estatísticas mais complexas [59].

Vários estudos que utilizam meta análise têm sido publicados na área de ciências médicas. O que faz a meta análise ser bem sucedida nesta disciplina científica é o fato de que os pesquisadores da área têm acesso a grandes bases de dados, construídos a partir de estudos experimentais realizados de forma consistente e padronizada. Esta padronização é refletida inclusive na escrita dos artigos que relatam estes estudos [16]. Estas bases de dados são as referências principais para identificar estudos existentes sobre um fenômeno específico, reduzindo assim a possibilidade de equívocos na divulgação dos resultados.

**Tabela 2 – Resumo de alguns métodos para síntese de evidências qualitativas e quantitativas**

<b>Método</b>	<b>Esboço da Abordagem</b>	<b>Problemas</b>	<b>Vantagens</b>
Sumário Narrativo	Descrição narrativa e ordenação de evidências primárias (talvez selecionadas) com comentário e interpretação.	Falta de transparência na seleção e também nos estágios posteriores do processo.	Procedimentos flexíveis; pode lidar com grandes bases e também com diversos tipos de evidência
Grounded theory	O método comparativo constante identifica padrões e inter-relacionamentos em dados primários; a amostragem responde à análise; os princípios da amostragem teórica podem ser usados	Falta de transparência	Busca por explicações/teorias generalizadas Podem incluir tanto evidências qualitativas quanto quantitativas
Análise qualitativa e comparativa	Análise booleana das condições necessárias e suficientes para que os resultados particulares sejam observados, baseadas na presença/ausência das variáveis e dos resultados em cada estudo primário	Focado na determinação da causalidade, aspectos não interpretativos dos dados qualitativos	Sistemático transparente; pode incorporar evidências qualitativas e quantitativas
Meta análise Bayesianas	A opinião quantificada sobre efeitos das variáveis dos estudos qualitativos formalmente combinadas (com o paradigma de Bayes) com a evidência dos estudos quantitativos	Conceitualmente simples mas pode ser tecnicamente complexo para executar (e pode perder assim a transparência)	O impacto da opinião prévia dos analistas pode explicitamente ser explorado
Revisões Sistemáticas	Busca detalhada, exaustiva por estudos primários a partir de uma questão de pesquisa específica, seleção dos estudos usando-se claramente critérios de elegibilidade reproduzíveis, avaliação crítica da qualidade dos estudos, e síntese dos resultados de acordo com um método pre-determinado e explícito. A análise é feita idealmente usando a Meta análise, mas as vezes também é feita usando outros métodos como contagem de votos.	Na maior parte das vezes é laborioso e custoso.	Captura precisamente a evidência.

<b>Método</b>	<b>Esboço da Abordagem</b>	<b>Problemas</b>	<b>Vantagens</b>
Meta análise	Combinação estatística dos dados por meio dos estudos para gerar a estimativa sumária dos efeitos. O termo “efeito” refere-se a qualquer medida de associação entre a exposição e o resultado (por exemplo, relação das probabilidades). Uma meta análise é geralmente a etapa final em uma revisão sistemática.	Não é sempre possível aplicar-se aos dados, os dados extraídos de estudos primários necessitam satisfazer certos requisitos (sendo o mais importante ter um alto nível de homogeneidade entre os estudos) Uma boa meta análise deve idealmente começar com uma revisão sistemática.	Pode ser feito sem uma revisão sistemática. Menos consumidor de tempo do que a revisão sistemática.
Contagem de Votos	Os resultados diferentes dos testes das hipóteses são categorizados como significativamente positivos, significativamente negativos ou sem efeito significativo. Cada estudo conta um “voto” na sustentação dos relacionamentos acima e os números dos votos são contados. Se a razão entre votos positivos e negativos for maior que um limite predeterminado, um relacionamento para a variável específica é identificado.	Não depende dos valores reais do tamanho do efeito e de medidas comparáveis. O método supõe que há um fenômeno comum subjacente, por exemplo, quando um único coeficiente de correlação é aplicado. Não considera a qualidade dos estudos.	É mais simples do que meta análise e pode ser aplicado tanto para estudos qualitativos quanto quantitativos.

---

Os pesquisadores em ES também têm aplicado meta análises em seus estudos [47, 51, 56]. A Seção 2.3.1 discute a aplicação da meta análise em ES bem como algumas de suas limitações para a área.

Uma abordagem interessante para a análise secundária em engenharia de software experimental é a contagem de votos, ou “vote counting”. Esta abordagem tenta acumular os resultados de uma coleção de estudos relevantes contando “quantos resultados são estatisticamente significativos em um sentido, quantos são neutros (isto é, “nenhum efeito”), e quantos são estatisticamente significativos no outro sentido” [123]. Esta abordagem é atraente para ES porque abstrai o conceito de resultado dos dados de baixo nível facilitando a análise de estudos heterogêneos. A Seção 2.3.2 discute a aplicação de contagem de votos em ES.

Pesquisadores de ES também estão usando revisões sistemáticas para combinar resultados dos estudos experimentais [14, 16, 65, 82, 92, 107, 126]. Uma revisão sistemática é uma maneira definida e metódica de identificar, avaliar e analisar estudos publicados a fim de investigar uma questão específica de pesquisa. Uma revisão sistemática pode também descobrir a estrutura e os padrões de certa pesquisa e assim identificar os pontos descobertos e que podem ser focados em pesquisas futuras [16]. As revisões sistemáticas diferem de revisões informais da literatura por serem formalmente planejadas e metodologicamente executadas. Uma boa revisão sistemática deve ser independentemente repetível e assim tem um valor científico muito maior do que um exame comum da literatura. O uso de revisões sistemáticas em ES tem aumentado nos últimos anos e já há vários estudos publicados sobre o assunto. A revisão sistemática em engenharia de software faz geralmente uso da meta análise tradicional como apoio à análise secundária [14, 16, 65, 82, 92, 107, 126]. A Seção 2.3.3 descreve o processo proposto por Kitchenham [16] que adapta os modelos tradicionais da revisão sistemática usados na área da medicina ao contexto da ES.

### **2.3.1 Meta análise em ESE**

Glass [48] define meta análise como “Uma análise estatística de uma grande coleção de análises, resultante de estudos primários com o propósito de integração dos resultados”. Ou

seja, a meta análise pode ser vista como uma integração estatística de estudos independentes sobre o mesmo tópico.

Mais formalmente, ela pode ser definida como o agrupamento estatístico dos dados por meio dos estudos para gerar o sumário dos efeitos das estimativas. O termo “efeito” refere-se a qualquer medida da associação entre a exposição e o resultado (por exemplo, relação das probabilidades). A maioria dos revisores começa a análise com a tabulação simples das características do estudo (por exemplo, o ano, o ajuste, o projeto do estudo, e a sua qualidade) e os seus resultados. Isto deve ser feito em todas as revisões sistemáticas, mesmo se nenhuma meta análise for ser executada. Diagramas de Forest (*Forest Plots*) mostram os efeitos de cada estudo com seus intervalos de confiança e fornecem um sumário visual dos dados. A etapa seguinte na análise é o agrupamento de medidas do efeito nos vários estudos. O agrupamento é essencialmente um processo de calcular médias ponderadas. Na meta análise, tipicamente, é atribuído mais peso para estudos maiores (isto é, aqueles estudos com tamanhos de amostra maiores e mais eventos). Quando existe heterogeneidade, pode não ser apropriado, por exemplo, combinar os efeitos de estudos observacionais com experimentos controlados [80].

Na meta análise pode-se adotar um modelo de efeito fixo ou um modelo aleatório. O modelo de efeitos fixos supõe que os estudos incluídos na meta análise estimam o mesmo efeito “verdadeiro” que é “fixo”, e que as diferenças observadas nos estudos são devidas ao erro aleatório (probabilidade). Por outro lado, o modelo de efeitos aleatórios pressupõe que os estudos incluídos na meta análise são somente uma amostra aleatória de um universo teórico de todos os estudos possíveis sobre uma dada questão de pesquisa, e que os efeitos para os estudos individuais variam em torno de algum efeito médio total.

Os modelos aleatórios incorporam duas fontes de variabilidade: o erro aleatório e a variabilidade entre-estudos. Conseqüentemente, o modelo aleatório dos efeitos é preferido quando os dados são heterogêneos, já que estes permitem a variabilidade entre-estudos e intra-estudo, e fornecem uma estimativa mais conservadora com intervalos maiores de

---

confiança. Na ausência de heterogeneidade, ambos os modelos produzem resultados similares [80].

A heterogeneidade refere-se a um grau elevado de variabilidade nos resultados dos estudos e não é incomum nas meta análises. Na presença de heterogeneidade significativa, a estimativa sumária não é significativa, já que é uma média de valores extremos e não descreve adequadamente os dados. De fato, os revisores podem escolher não forçar os resultados em uma única estimativa sumária na presença de heterogeneidade. Neste caso, os revisores devem se focar preferivelmente em encontrar fontes potenciais de estimativas de variabilidade. Isto pode ser realizado por métodos tais como análises do subgrupo, meta-regressão e métodos gráficos [80].

Os primeiros estudos de meta análise surgiram nas áreas de psicologia e educação, na década de 70. Vários pesquisadores destas áreas aplicaram a meta análise estatística em seus estudos [47, 51, 56].

A aplicação de meta análises tem sido bem mais modesta na área de engenharia de software experimental. A Tabela 3 descreve alguns artigos que aplicaram a meta análise em estudos experimentais em ES. Como podemos perceber, além de poucos trabalhos na área, a maior parte deles demonstra a inabilidade da aplicação de meta análise para os propósitos das suas análises. Existem, basicamente, três problemas recorrentes relacionados à aplicação da meta análise à ES:

- Ela requer que os dados estejam disponíveis para a análise, mas isto nem sempre é verdade para os estudos em ES;
- Ela geralmente necessita que os estudos tenham as mesmas hipóteses iniciais, o que nem sempre ocorre em ES;
- Ela necessita que os protocolos de pesquisa prevejam o relato de informações sobre resíduos (erros) das análises, o que não é comum em ES.

**Tabela 3 - Meta análise em ES.**

Referência	Abordagem	Estudo Experimental	Conclusões principais na aplicação da Meta análise
Porter e Johnson - 1997 [2]	Reconciliação (compreendem-se as similaridades e as diferenças entre dois estudos de modo que seus resultados possam ser comparados).	Uma análise comparativa detalhada de dois experimentos. Os experimentos tiveram uma motivação similar: avaliar as contribuições reais das reuniões para revisão de software. Diferiram em seus projetos os instrumentos, os documentos para revisão, procedimentos, e medição.	<ul style="list-style-type: none"> <li>• O processo de reconciliação destacou muitas similaridades e diferenças entre os dois experimentos, permitindo a melhor compreensão de quais dados eram comparáveis e quais não eram.</li> <li>• O processo de reconciliação insere um contrapeso entre as limitações de uma comparação baseado na literatura e as restrições da meta análise.</li> <li>• Os experimentos diferiram em muitas maneiras e os autores ainda são céticos quanto ao fato de que a meta análise tradicional poderia ser aplicada com sucesso nesses estudos.</li> <li>• A reconciliação permitiu executar uma comparação das hipóteses e dos resultados, e observar similaridades entre os resultados combinados.</li> </ul>
Pickard e outros - 1998 - [75]	Meta análise e contagem de votos. A contagem de votos consiste em categorizar os resultados diferentes dos testes de hipóteses em três grupos: a) Significativo efeito positivo; b) Significativo efeito negativo; c) Efeito não-significativo;	Um exemplo de estimativas do custo do software. Eles investigaram o relacionamento entre o esforço do projeto e o tamanho do produto encontrado em diferentes estudos em ES. Os conjuntos de dados são provenientes de estudos observacionais e não de experimentos formais.	<ul style="list-style-type: none"> <li>• A investigação mostrou que há uma correlação elevada entre os estudos e que a heterogeneidade dos dados não influencia o resultado na maior parte dos casos.</li> <li>• A contagem de votos tem problemas inerentes que indicam que não deve ser usada como um método para combinar resultados de estudos experimentais.</li> <li>• A meta análise permite que os pesquisadores avaliem o efeito do tamanho estimado a partir dos efeitos de tamanho de cada estudo individual, e levá-los a inferências estatísticas muito mais fortes do que poderia ser feito pela contagem de votos.</li> </ul>

Referência	Abordagem	Estudo Experimental	Conclusões principais na aplicação da Meta análise
Hayes - 1999 - [137]	Meta análise usando testes combinados de valores de significância e do efeito das magnitudes.	Cinco experimentos publicados na literatura foram analisados. O objetivo do experimento original era estudar o método de inspeção em ES. Desde a publicação do primeiro experimento, quatro replicações foram relatadas com diferentes resultados.	<ul style="list-style-type: none"> <li>• O artigo não fornece uma decisão final sobre a possibilidade de adoção da abordagem, mas fornece uma discussão sobre problemas associados com a heterogeneidade presente nos conjuntos de dados para este tipo de análise.</li> </ul>
Miller - 2000 [59]	Meta análise usando efeitos do tamanho, testes difusos e testes focados.	Este artigo analisa experimentos baseados em hipóteses similares. Cinco estudos independentes foram analisados, nos quais a premissa básica investigada por estes experimentos é: “Qual técnica de detecção de defeitos é mais efetiva na descoberta de defeitos?”.	<ul style="list-style-type: none"> <li>• O autor afirma que não foi possível resolver a natureza heterogênea do conjunto de dados; assim, não pode continuar a análise e combinar os dados.</li> <li>• O autor afirma que a disciplina de ES deve iniciar um período de melhoria para reduzir a variabilidade entre os experimentos replicados ou os experimentos que examinam a mesma hipótese.</li> </ul>
Djokicf - 2001 [118]	Estimativas com peso e correlação comum.	Projetos de domínio público que foram coletados na web.	<ul style="list-style-type: none"> <li>• O teste de homogeneidade é satisfeito para três dos casos;</li> <li>• O resultado foi considerado bom pelos autores.</li> </ul>
Runeson e Thelin - 2003 - [109]	Análise de Validação Cruzada. Os testes estatísticos utilizados foram: Kruskal-Wallis e o procedimento de comparação múltipla para testes não-paramétricos.	Quatro experimentos compararam técnicas de leitura baseadas em uso (UBR – Usage –Based Reading) a quatro outras técnicas alternativas.	<ul style="list-style-type: none"> <li>• Para os autores, a meta análise oferece a oportunidade de analisar resultados de experimentos diferentes com as mesmas hipóteses. Entretanto, isto é executado sob um determinado risco se os ajustes experimentais não forem bem similares ou simplesmente insuficientemente detalhados para permitir comparações.</li> </ul>

### **2.3.2 Contagem de Votos (Vote Counting)**

Esta abordagem visa acumular os resultados de uma coleção de estudos relevantes contando “quantos resultados são estatisticamente significativos em um sentido, quantos são neutros (isto é, “nenhum efeito”), e quantos são estatisticamente significativos no outro sentido” [123]. A categoria que tem a maioria de contagens, ou votos, é selecionada para representar o resultado típico, indicando desse modo os meios mais eficazes da intervenção.

Um problema das revisões de contagem de votos é que elas não consideram o fato de que alguns estudos são metodologicamente superiores a outros e, conseqüentemente, merecem um peso especial na análise [123]. Um outro problema com a contagem de votos é que ela não indica “a possibilidade de que um tratamento poderia ter tido conseqüências diferentes sob circunstâncias diferentes” [123].

Mohagheghi e Conradi aplicaram a contagem de votos em resultados quantitativos em Engenharia de Software e afirmam que o método é aplicável quando outros métodos tais como a meta análise não podem ser aplicados [105]. Outros autores como Pickard et. al. [75] apontam que enquanto meta análise e contagem de votos são dois métodos que podem ser usados para análise de resultados quantitativos, uma das vantagens de contagem de votos é que ela também pode ser usada em estudos qualitativos.

A contagem do voto é mais simples do que a meta análise e é recomendada quando os estudos usam medidas incompatíveis (desde que satisfaçam critérios mínimos de qualidade) ou a informação relatada nos estudos é muito limitada [105].

### **2.3.3 Revisão Sistemática em ES**

Uma revisão sistemática corresponde a uma abordagem de revisão de literatura que visa estabelecer um processo formal para esse tipo de investigação. Ela tenta evitar as principais limitações de uma revisão informal como, por exemplo, a introdução de eventuais vieses na escolha de documentos e extração de informações, ou a dificuldade de repetição do estudo pela falta de documentação do seu protocolo de execução. A revisão sistemática atua como um meio para identificar, avaliar e interpretar toda pesquisa disponível e relevante sobre uma questão de pesquisa, um tópico ou um fenômeno de interesse. A condução de uma revisão

---

sistemática supostamente apresenta uma avaliação justa do tópico de pesquisa, na medida que utiliza uma metodologia de revisão rigorosa, confiável e passível de auditoria [16].

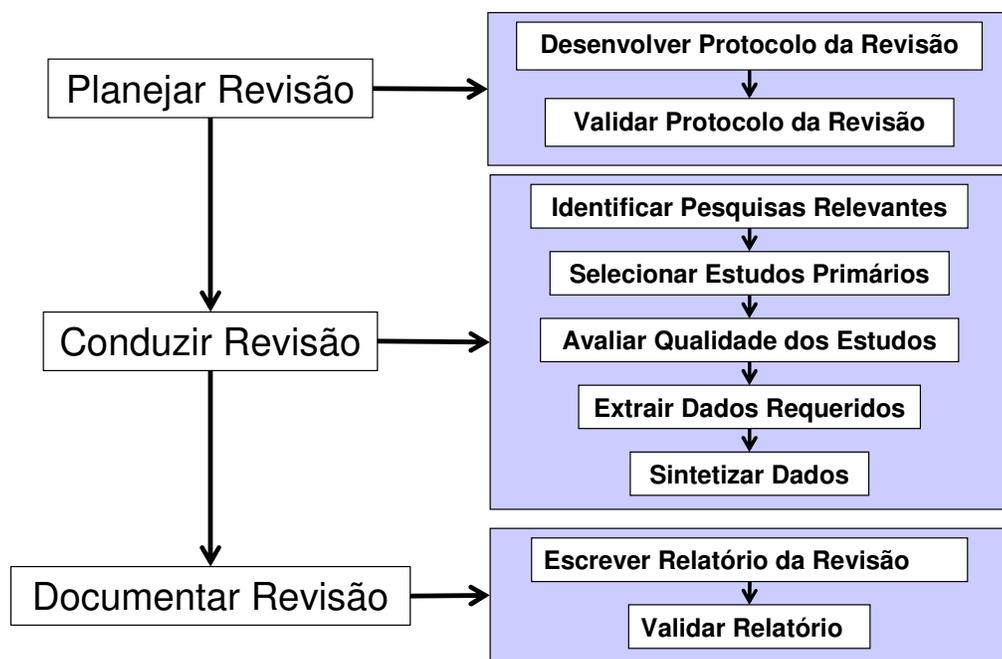
Há muitas razões para se realizar uma revisão sistemática. As razões mais comuns segundo Kitchenham são:

- Resumir a evidência existente a respeito de um tratamento ou de uma tecnologia; por exemplo, para resumir as evidências experimentais dos benefícios e das limitações de um método específico;
- Identificar algumas lacunas na pesquisa atual a fim de sugerir áreas para investigações futuras;
- Fornecer um arcabouço a fim de posicionar apropriadamente atividades novas de pesquisa.

As seguintes características distinguem uma revisão sistemática de uma revisão convencional de literatura [16]:

- A definição e a documentação de um protocolo para a revisão sistemática antes da execução da mesma, com o objetivo de se especificar explicitamente as questões de pesquisa e os procedimentos a serem usados no estudo;
- Definição e documentação de uma estratégia de busca como parte do protocolo, para certificar de forma auditável que foram encontrados os estudos relevantes da literatura;
- Descrição, como parte do protocolo, de critérios explícitos para inclusão e exclusão dos estudos que irão fazer parte da revisão;
- Descrição, como parte do protocolo, dos mecanismos de avaliação da qualidade dos estudos a serem analisados;
- Descrição, como parte do protocolo, dos processos de revisão e de verificação cruzada, envolvendo múltiplos pesquisadores independentes, a fim de controlar os erros individuais de cada pesquisador.

Kitchenham [16] propõe algumas diretrizes para revisões sistemáticas em ES. Ela descreve as três fases principais de um processo sistemático de revisão: planejar a revisão, conduzir a revisão e escrever o relatório da revisão, como mostrado na Figura 7. Cada uma destas fases contém uma seqüência de atividades, mas a execução do processo como um todo envolve iterações e refinamentos do processo formalmente definido. É importante notar que esta abordagem direciona para o uso da meta análise na sintetização dos dados na condução da revisão.



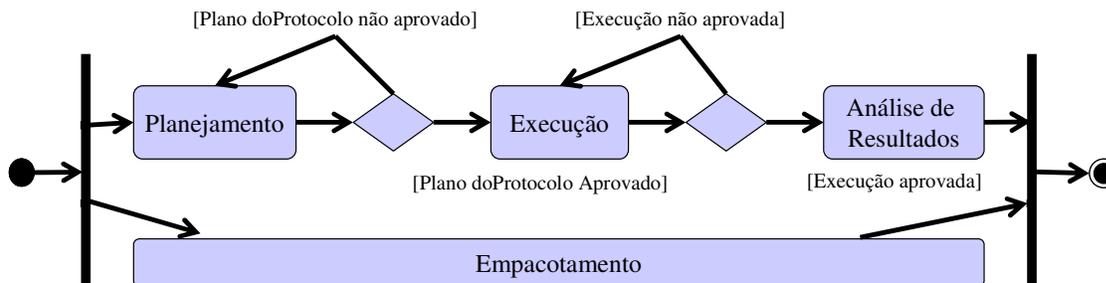
**Figura 7 - Abordagem de Kitchenham para Revisão Sistemática**

Baseados na abordagem de Kitchenham, Biolchini e outros definiram um processo mais detalhado para condução da revisão sistemática em ES [65]. Este processo é mostrado na Figura 8. Durante a sua fase de planejamento, os objetivos da pesquisa são listados e um protocolo de revisão é definido. Tal protocolo especifica a questão central da pesquisa e os métodos que serão usados para executar a revisão. O estágio da execução envolve estudos preliminares nos artigos, identificação, seleção e avaliação dos artigos de acordo com os critérios de exclusão e inclusão estabelecidos no protocolo da revisão. Uma vez que os

---

estudos foram selecionados, os dados dos artigos são extraídos e sintetizados durante a fase de análise dos resultados. Seus resultados devem ser armazenados à medida que cada uma destas fases é executada.

Os protocolos de trabalho e toda decisão, informação obtida, lição aprendida, ou artefato produzido deve ser empacotado para reuso futuro. O empacotamento é executado durante todo o processo de revisão. Há pontos de verificação no processo sistemático proposto. Antes de executar a revisão sistemática, é necessário garantir que o planejamento está apropriado. O protocolo deve ser avaliado e se problemas forem encontrados, o pesquisador deve retornar à fase de planejamento para rever o protocolo. Similarmente, se problemas a respeito dos mecanismos de busca forem encontrados durante a fase da execução, a revisão sistemática pode ser re-executada. Kitchenham não fala explicitamente sobre a fase de empacotamento, sendo esta uma das principais diferenças entre as abordagens em termos de processo.



**Figura 8 – Abordagem de Biolchini e outros para Revisão Sistemática**

A realização de uma busca sistemática é uma das principais diferenças entre uma revisão tradicional e uma revisão sistemática da literatura. O alvo é identificar o conjunto de estudos mais completo sobre o tópico de interesse. Concordamos com Davis e outros [4] quando estes afirmam que nenhum método especial parece ser necessário a priori, para selecionar artigos em engenharia de software experimental, porque é improvável que existam muitos estudos experimentais sobre um determinado tópico. Entretanto, ao realizar uma busca para revisão sistemática, é importante que o revisor mantenha em mente que:

- 1) Não é fácil encontrar artigos experimentais de interesse, porque não existem muitos e também existe uma grande variedade de descritores usados na sua indexação em bases de dados bibliográficas;
- 2) Os tipos de estudos experimentais usados em ES (experimentos, pesquisas de opinião, estudos de caso, etc.) podem diferir extremamente de um artigo para outro;
- 3) O projeto, os fatores e as variáveis de resposta usadas nos estudos são muitos;
- 4) Independentemente da metodologia usada, a execução dos estudos experimentais pode ainda conter ameaças à validade dos resultados;

Davis e outros [4] também afirmam que, além de tudo, não é fácil agregar toda a evidência recolhida de maneira a assegurar que algumas evidências não sejam desprezadas ou que outros problemas de análise não ponham em risco a validade da revisão. Vale a pena mencionar que não ter um método específico ou o procedimento estabelecido para empreender esta agregação afeta diretamente a repetibilidade dos resultados da revisão.

Um aspecto importante na revisão sistemática é a avaliação da qualidade dos estudos que serão revisados. Kitchenham [16] afirma que uma dificuldade inicial para a avaliação da qualidade é que não há nenhuma definição estabelecida de “qualidade” para estudos experimentais em engenharia de software. Entretanto, as diretrizes do Centre for Reviews and Dissemination CRD [74] e o manual dos revisores de Cochrane [30] sugerem que a qualidade se relaciona à extensão na qual o estudo minimiza os equívocos e maximiza a validade interna e externa da análise. As principais categorias de riscos associados à validade de estudos experimentais compiladas por Kitchenham a partir destas diretrizes são:

- 1) Equívocos ou Erros Sistemáticos: A tendência a produzir resultados que diferem sistematicamente dos resultados “verdadeiros”.
- 2) Validade Interna: A extensão na qual o projeto e a condução do estudo são prováveis de prevenir o erro sistemático. Validade Interna é um pré-requisito para a validade externa.
- 3) Validade Externa ou Aplicabilidade: A extensão nas quais os efeitos observados no estudo são aplicáveis fora do contexto do estudo.

---

Não é nosso foco destacar as diferenças e as similaridades das diversas abordagens de revisão sistemática. Biolchini e outros fazem um trabalho muito bom neste sentido, com referências a outras abordagens até mesmo fora da área de ES [65, 66]. Há alguns artigos sobre experiências na condução de revisões sistemáticas baseadas nas diretrizes de Kitchenham [14, 82, 92]. Staples e outros [92] comentam as suas experiências na condução de revisões sistemáticas usando as diretrizes de Kitchenham e indicam a necessidade da inclusão de melhor orientação sobre como criar os protocolos durante seu desenvolvimento.

Mian e outros [106] propõem o uso de um modelo para um processo de revisão sistemática para orientar o pesquisador de ES. O modelo e o processo foram usados pela equipe de Engenharia de Software Experimental da COPPE/UFRJ para executar revisões sistemáticas em diferentes tópicos de ES [65]. Jorgensen e outros [82] conduziram uma revisão sistemática de estudos de estimativas de custo no desenvolvimento do software e, baseados na revisão, fornecem recomendações para pesquisas futuras em estimativas de custo do software.

Apesar da existência de relatos de experiências de aplicações de revisões sistemáticas em áreas do conhecimento onde ela está bem estabelecida, e das recentes aplicações em ES, discutidas acima, muito trabalho ainda é necessário para estabelecer uma abordagem eficiente para a área de ES. Desenvolver esta abordagem de investigação no campo da engenharia de software implica adaptar as dimensões conceituais e metodológicas da síntese da pesquisa ao domínio, considerando suas especificidades como uma área de conhecimento científico. Diferentemente da área médica, a engenharia de software tem algumas especificidades que dificultam a síntese da pesquisa para obtenção de evidências [66]:

- A maioria dos métodos e técnicas da engenharia de software devem ser executados por profissionais experientes, que conheçam bem não somente as práticas de sua aplicação, mas também como eles funcionam. Já em medicina, embora os médicos sejam qualificados na prática de sua profissão, eles não precisam necessariamente saber como atua o tratamento prescrito (por exemplo, como funciona um remédio) para aplicá-lo adequadamente..

- A maioria das técnicas de ES afetam o ciclo de vida do software tornando difícil isolar o efeito individual de uma técnica. As técnicas de interesse normalmente interagem com muitas outras técnicas e procedimentos do desenvolvimento. No geral, é difícil determinar uma relação causal linear entre uma técnica particular e um resultado desejado do projeto.
- Diferentemente da engenharia de software, os investigadores da área médica dispõem de muitas revisões sistemáticas já publicadas, isto é, artigos que trazem todos os relatórios relevantes de um tópico em particular. Os pesquisadores médicos têm uma boa infra-estrutura tecnológica e científica para apoiá-los nesta tarefa. Há diversas organizações (por exemplo, a colaboração internacional Cochrane - [www.cochrane.com](http://www.cochrane.com)) que montam revisões sistemáticas dos estudos sobre drogas e procedimentos médicos. Para fornecer uma fonte de informação central para a evidência, o Cochrane torna públicas estas revisões sistemáticas que são revisadas continuamente. Atualmente, ainda não há nada equivalente à colaboração Cochrane na área de Engenharia de Software.

Além das diferenças que dificultam o uso do processo de revisão sistemática na Engenharia de Software, uma das questões também apontadas por Biolchini et. al. [66] é que nas suas experiências na aplicação de revisões sistemáticas na área (mais de 5), não foi possível aplicar a meta análise formal dos resultados. O problema mais freqüente associado a esta dificuldade é a não inclusão das informações sobre resíduos (erros) das análises no artigos analisados.

## **2.4 Limitações e Vantagens dos Métodos de Análise Secundária**

Sjoberg e outros [32] fizeram uma revisão da literatura sobre experimentos controlados na área de ES. Sua finalidade era caracterizar os tópicos dos experimentos, seus participantes (número de participantes, tipos de participantes, o recrutamento, as recompensas para a participação), e as tarefas executadas durante o experimento (tipo de tarefa, duração, tipo e tamanho da aplicação). Além disso, este artigo relata como as validades interna e externa são

---

tratadas e a extensão na qual os experimentos são repetidos. Todos os artigos eram de 1993 a 2002, de diferentes periódicos em ES. Os autores avaliaram 5.453 artigos e analisaram 103 artigos em profundidade. O principal resultado deste trabalho é que os artigos em engenharia de software experimental são geralmente vagos e não sistemáticos e que não existe uma terminologia consistente nos relatos dos experimentos.

Mohaguegui e Conradi [105] levantam uma pergunta similar: Temos exemplos sobre a combinação de evidências quantitativas na pesquisa de engenharia de software experimental. Para eles, a resposta até agora é não. Um diagnóstico com o qual nós concordamos. Para ilustrar a situação, considere as três revisões sistemáticas publicadas na IEEE Transactions on Software Engineering em 2007 [14, 32, 82]. Jørgensen e Shepperd [82] fornecem uma base para a melhoria da pesquisa de estimativas de software com uma revisão sistemática sobre trabalhos precedentes. Eles identificaram 304 artigos de estimativas de custo de software em 76 periódicos e classificaram os artigos de acordo com o tópico da pesquisa, a abordagem da estimativa, a abordagem de pesquisa, o contexto do estudo e o conjunto de dados. Sjoberg e outros [32] relatam uma revisão sistemática em um conjunto de 103 artigos que relatam estudos experimentais, de um total de 5.453 artigos publicados em periódicos principais e em conferências da Engenharia de Software na década 1993-2002, sua finalidade era caracterizar os tópicos dos experimentos, seus participantes (número de participantes, tipos de participantes, o recrutamento, as recompensas para a participação), e as tarefas executadas durante o experimento (tipo de tarefa, duração, tipo e tamanho da aplicação). Kitchenham e outros [14] executaram uma revisão sistemática dos estudos comparando modelos de predições coletadas dentro da própria organização com os modelos coletados fora da organização. Estes trabalhos revisam distribuições das amostras, médias, medianas e desvios padrões, mas não combinam os resultados quantitativos nos tópicos específicos. Em outras palavras nenhum deles aplica meta análise dos resultados.

A partir desta discussão, pode-se perceber que tem sido difícil combinar quantitativamente os resultados de estudos primários em engenharia de software. Desta forma, abordagens baseadas em contagem de votos aparecem como opções práticas e viáveis no cenário atual.

Não se pode porém, descartar a preocupação da revisão sistemática com a formalização do processo de análise e com a qualidade dos estudos primários sob investigação. Esta é uma das maiores contribuições da metodologia para a área de estudos secundários. Por exemplo, as revisões sistemáticas tentam distinguir entre os estudos de maior e de menor tamanho e definem pesos específicos para cada um de acordo com este fator. O método também direciona para a incorporação de alguns fatores contextuais analisando os resultados e limitações dos diversos estudos, identificando suas implicações para a prática do tópico em questão. Este nível de detalhe e rigor não tem sido tradicionalmente usado em metodologias baseadas em contagem de votos.

Nosso trabalho se foca exatamente em combinar o que há de melhor em contagem de votos e revisões sistemáticas para propor uma abordagem para realização de estudos secundários especificamente voltada para a área de engenharia de software. Denominamos esta abordagem de SecESE (Análise Secundária em Engenharia de Software Experimental). Ela combina a simplicidade e flexibilidade de análise de resultados usada em contagem de votos com a formalidade e aspectos de formalização que é usada em revisão sistemática.

A SecESE propõe que a análise de estudos deve ser feita individualmente e a “meta análise” deve ser feita no âmbito de resultados obtidos a partir da análise primária, não no âmbito da análise de dados. Em vez de trabalhar em um nível baixo de abstração – dados, focamos no nível do conhecimento extraído destes dados – análise de resultados.

A SecESE tem duas fases principais: a extração e a análise dos resultados. A primeira extrai a informação dos dados ou de relatórios (artigos). A segunda analisa estes resultados por meio de uma abordagem baseada em mineração visual de dados.

A SecESE usa aspectos de formalização que são encontrados nas revisões sistemáticas. Na definição dos objetivos do estudo secundário, ela usa uma abordagem baseada no GQM [133] para estabelecer e detalhar os objetivos da análise. Na seleção de estudos, ela usa critérios similares aos usados na revisão sistemática. E na extração de informações, ela define uma abordagem especificamente voltada para a área de engenharia de software experimental, coletando não só os resultados, mas também as informações de contexto dos estudos

---

primários. A qualidade dos estudos também é considerada através de uma abordagem baseada em pontuação quantitativa dos estudos primários.

A informação extraída (resultados, contexto e qualidade) é armazenada em uma base de dados, onde esta pode ser analisada para extrair conclusões. O processo de análise dos resultados é também baseado em exploração visual e mineração de dados. Ela define um método específico de contagem de votos apoiado por técnicas de visualização e aglomeração de informações. A abordagem SecESE é descrita em detalhe no próximo capítulo.

## Capítulo 3

### Abordagem para Abstração de Conhecimento a partir de Estudos Primários (SecESE)

#### 3.1 Definições e Visão Geral

A Engenharia de Software Experimental é baseada na coleta de dados, compreensão, extração da informação e abstração do conhecimento sobre quando e como os métodos e as técnicas podem e devem ser aplicados no desenvolvimento e manutenção do software.

Uma vez que o objetivo principal da SecESE é extrair conhecimento de informação experimental necessitamos trazer um pouco desta discussão para o contexto deste trabalho. Como não há definição comum para “conhecimento” e de como este conceito é diferente de dados, informação e sabedoria, o que se segue são as definições que usaremos para estes termos, e como elas são usadas para aumentar a compreensão de um determinado assunto [103, 116]. A Figura 9 descreve graficamente estes conceitos e suas relações [116].

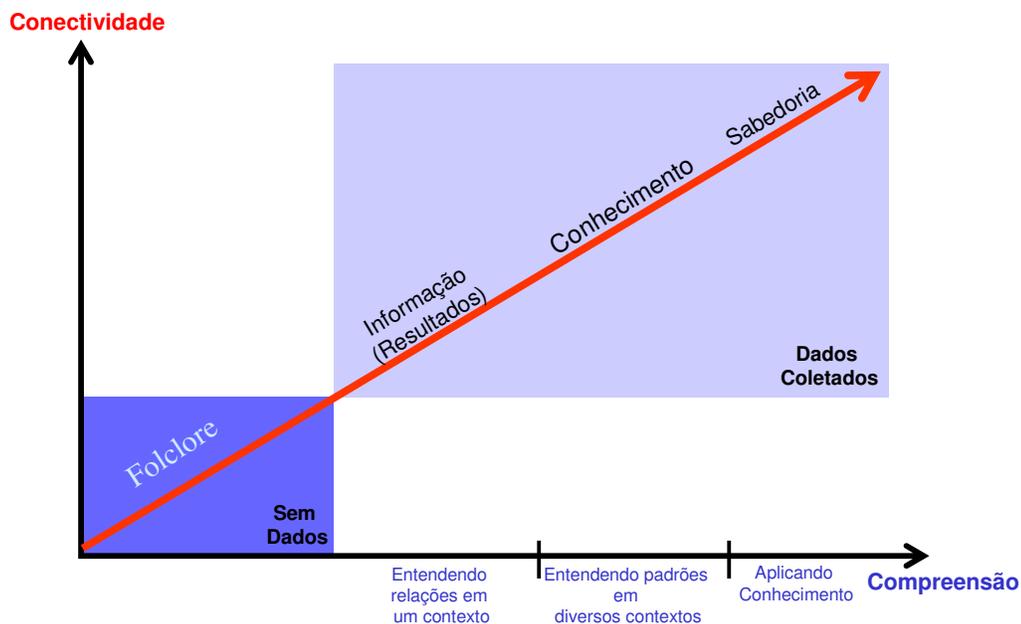


Figura 9 - Dados, Informações e Conhecimentos com relação à Compreensão e Conectividade.

---

**Folclore** é um conjunto de mitos, crenças, histórias populares, lendas, tradições e costumes que são transmitidos de geração para geração e que fazem parte da cultura popular [79]. A palavra folclore vem do inglês “folk” = povo e “lore” = conhecimento e significa sabedoria popular.

Para os cientistas, enquanto não há dados experimentais para apoiar uma afirmação não há evidência de que a afirmação seja confiável. A experimentação é o centro do processo científico. Ela oferece um modo sistemático, disciplinado, computável e controlado para a avaliação de fenômenos naturais e sociais.

Os objetivos dos estudos experimentais na Engenharia de Software são a caracterização, avaliação, previsão, controle e melhoria de produtos, processos, recursos, teorias e modelos associados ao desenvolvimento e à manutenção de software [53]. Medição é a parte central de um estudo experimental. É definida como o mapeamento do mundo experimental para o mundo formal. O objetivo principal da medição na engenharia de software é aumentar a compreensão do processo e do produto para controlá-los, definindo atividades preventivas e corretivas, identificando suas possíveis áreas de melhoria [53].

Neste cenário, **dados** resultam essencialmente da coleta de medidas. Os dados não têm significado a menos que se compreenda o contexto em que foram coletados. Quando os dados são compreendidos associados a um contexto tem-se então uma informação. Os dados são transformados em informação a partir de sua organização de forma a identificar o que eles significam e em que condições foram obtidos. Neste trabalho chamaremos esta informação de **informação bruta** ou **informação crua**.

No contexto da SecESE, a análise da informação bruta gera os resultados ou **evidências**, que são as **informações codificadas** quase sempre referenciadas aqui como informação. Um resultado é uma tentativa de explanação para determinados comportamentos, fenômenos, ou eventos que ocorreram em um estudo experimental. Na SecESE, a informação codificada é obtida por meio dos resultados ou evidências que são encontrados nos documentos científicos. Estes termos serão usados como sinônimos no restante deste trabalho.

A informação codificada transforma-se em **conhecimento individual** quando é aceita e retida por um indivíduo como sendo uma compreensão apropriada do que é verdadeiro, representando o que este indivíduo considera uma interpretação válida da realidade. Similarmente, o **conhecimento comum** existe quando é aceito por consenso de um grupo de pessoas. O conhecimento comum não requer necessariamente que seja compartilhado por todos os membros de uma comunidade para existir. O fato de que ele é aceito em um grupo de pessoas pode ser considerado uma condição suficiente. Entretanto, é importante notar que o fato de estar disponível em um material publicado não dá a uma informação a condição de “conhecimento comum”.

A **sabedoria** é a habilidade de aplicar o conhecimento, experiência, compreensão ou senso comum. No exemplo da Figura 10, a sabedoria se manifesta quando o usuário da informação sabe quando e como aplicar as técnicas em um contexto específico.

DADOS			
S	2	35	A
S	2	30	B
S	0	20	B
S	1	37	A
N	1	24	A
N	2	17	A
N	0	8	B
N	2	18	B

Informação Crua			
Usou Técnica?	Experiência	N. de Defeitos Detectados	Produto de Trabalho
S	2	35	A
S	2	30	B
S	0	20	B
S	1	37	A
N	1	24	A
N	2	17	A
N	0	8	B
N	2	18	B

Evidências ou Resultados
A efetividade em detecção de defeitos é maior para participantes aplicando a técnica
A efetividade em detecção de defeitos é maior para participantes experientes

**Figura 10 - Exemplos de Dados, Informações Brutas e Resultados**

Na ESE, os dados são coletados durante um estudo experimental. Quando o analista considera seu contexto estes dados transformam-se em informações brutas; quando ele reconhece relações existentes nestas informações elas transformam-se em informações codificadas (evidências ou resultados). Quando existem resultados sobre os quais o analista é

---

capaz de reconhecer padrões ou realizar abstrações e generalizações em um determinado contexto, obtém-se o que chamamos de conhecimento. Fora do escopo da análise, chamamos de sabedoria o fato de alguém saber tirar proveito do conhecimento obtido.

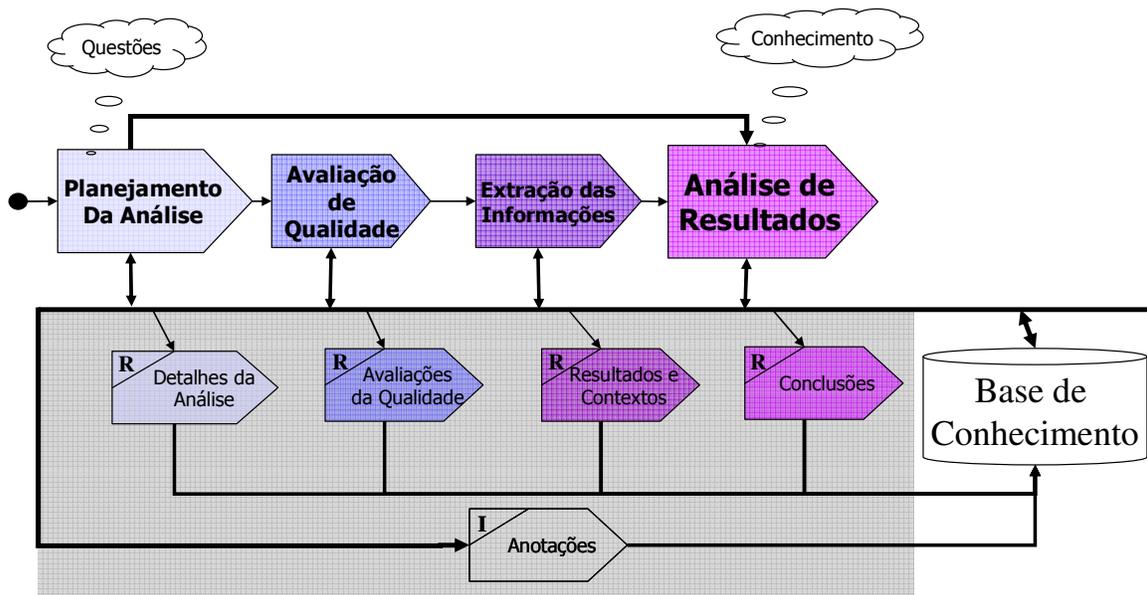
A SecESE é baseada na teoria de que podemos combinar resultados dos estudos primários diversos quando analisamos a informação codificada, considerando o contexto dentro do qual estes resultados foram coletados e analisados. Isto permite combinar a informação de muitos estudos primários. Propomos que a análise de cada estudo primário deve ser feita individualmente (análise primária) e que a análise secundária seja feita diretamente sobre a informação codificada, sem combinar a informação bruta contida em cada estudo.

Esta proposta de abstração é a principal contribuição da abordagem SecESE. Ela permite que análises secundárias em engenharia de software possam combinar resultados obtidos de fontes heterogêneas de informação bruta. Ela é inclusiva em relação a fontes de informações em que os dados ou a informação bruta não estão disponíveis e, é capaz de lidar com a falta de padronização dos documentos científicos produzidos na área.

### **3.2 A Abordagem SecESE**

Como mostrado na Figura 1 do Capítulo 1, abordagem SecESE é dividida em quatro atividades principais: planejamento, avaliação de qualidade, extração e análise de resultados. Para a primeira atividade existe a necessidade de definição e planejamento da análise e da explicitação dos critérios de busca e seleção dos objetos de análise. Após a seleção dos objetos de análise, deve-se então avaliar a qualidade dos mesmos. Para a atividade de extração só serão considerados os objetos de análise que tiverem um nível aceitável de qualidade de acordo com os critérios estabelecidos pelo analista; as grandes fontes de informação experimental (objetos de análise) podem ser os dados ou relatórios do estudo experimental. Como realçado anteriormente, o foco deste trabalho tem como objeto de análise, os relatórios experimentais. Na atividade de extração são realizadas as extrações das informações codificadas dos artigos. Na quarta atividade é realizada a análise das informações coletadas durante todo o processo.

A Figura 11 detalha a abordagem SecESE mostrando as fases de planejamento da análise e de avaliação da qualidade da informação além das fases de extração e de análise de resultados. A notação usada na figura é a proposta por Villela [71]. A SecESE começa com as questões de pesquisa, passa pelo planejamento da análise, a extração dos resultados primários, a avaliação da qualidade da informação coletada e a análise dos resultados, finalizando o processo quando o conhecimento é adquirido.



**Figura 11 - Abordagem SecESE para Análise de Resultados em ES.**

As atividades Planejamento, Extração, Avaliação de Qualidade e Análise Secundária são as principais do processo, e geram as atividades de registro na base de dados:

- Na atividade de Planejamento, os detalhes da análise são registrados na base de conhecimento;
- Na avaliação da qualidade, as avaliações dos documentos;
- Na extração, os resultados e informações de contexto coletadas dos documentos experimentais;
- Na análise de resultados são coletadas as conclusões do processo que são registradas na base.

- Em todas as atividades algumas anotações são necessárias para que a repetibilidade do processo seja possível; por isto existe a atividade implícita em todo o processo de registrar as Anotações.

As atividades principais da SecESE são descritas na Tabela 4 a seguir.

**Tabela 4 - Macro atividades da SecESE**

<b>Atividade: Planejamento da Análise</b>	
Objetivo	Definir alguns detalhes da análise. No início definem-se os objetivos da análise. Cada pergunta da análise pode originar um ou mais objetivo de análise. Os artigos são selecionados baseados nestes objetivos. E então os parâmetros iniciais, limitações, expectativas, metodologia aplicada na análise são indicados.
Entradas	Questões de Pesquisa
Subatividades	Definir objetivos de Análise Selecionar Artigos Definir Metodologia de Análise
Produtos	Plano de Análise
Responsáveis	Especialista em Análise
<b>Atividade: Avaliação de Qualidade</b>	
Objetivo	Avaliar detalhadamente a qualidade dos artigos. As avaliações de qualidade são baseadas em uma lista de verificação dos fatores que necessitam ser avaliados para cada estudo.
Entradas	Artigos Selecionados
Subatividades	Avaliar Artigos
Produtos	Artigos Avaliados
Responsáveis	Especialista em Análise
<b>Atividade: Extração de Informações</b>	
Objetivo	Extrair a informação dos artigos sob análise.
Entradas	Artigos Avaliados
Subatividades	Extrair Resultados Extrair Informações de Contexto
Produtos	Resultados de Cada Artigo e Informações de Contexto associados aos resultados.
Responsáveis	Especialista no Domínio e Especialista em Experimentação
<b>Atividade: Análise de Resultados</b>	
Objetivo	Analisar os resultados obtidos dos artigos conjuntamente com as informações de contexto dos mesmos.
Entradas	Informações Extraídas dos Artigos
Subatividades	Formalização dos Resultados Interpretação dos Resultados
Produtos	1) Um conjunto de conclusões ou conhecimento; 2) Uma base estruturada de resultados e informações de contexto; 3) Conhecimento tácito do processo (anotações, decisões, lições aprendidas, resultados intermediários).
Responsáveis	Especialista no Domínio, Especialista em Experimentação e Especialista em Análise.

É importante observar que existem três papéis principais nesta abordagem: o analista dos dados, que é a pessoa que executará a análise, extraindo as informações dos relatórios e organizando os resultados; o especialista do domínio, que é o fornecedor das questões de pesquisa e quem interpreta os resultados em última instância; e o especialista em experimentação, que é o profissional responsável por verificar a validade e o escopo dos resultados obtidos na análise. As seções seguintes explicam cada uma das atividades da SecESE em detalhe.

### **3.3 Planejamento da Análise**

O processo começa com perguntas gerais que um analista quer investigar em estudos experimentais. É importante destacar que a SecESE pressupõe que o analista está trabalhando com estudos já existentes. O objetivo principal da SecESE é usar os resultados que já existem para extrair a informação e produzir o conhecimento novo num contexto mais amplo, as origens das questões podem ser:

- Folclores: A literatura “não escrita” da cultura em Engenharia de Software gera algumas questões para os pesquisadores interessados em saber se elas são válidas ou não.
- Hipóteses: a literatura escrita em Engenharia de Software gera algumas questões que o pesquisador quer investigar.
- Questões explicitamente estabelecidas em um estudo experimental: todas as questões do projeto experimental são questões que o pesquisador pode querer investigar em outros contextos.
- Objetivos mais gerais de um projeto (ou estudo): estas questões muitas vezes não foram previstas quando um estudo experimental foi projetado, mas surgem na análise.

Kitchenham [16] diz que um ponto crucial é saber fazer as perguntas certas. Estas perguntas são geralmente as que:

- 
- Têm um significado importante tanto para os pesquisadores quanto para os profissionais. Por exemplo, os pesquisadores podem estar interessados em saber se uma técnica específica de análise conduz a uma estimativa mais exata dos defeitos restantes após inspeções no projeto. Entretanto, um profissional pode querer saber se adotar uma técnica específica de análise para prever defeitos remanescentes é mais eficaz do que a opinião de um perito em identificar os documentos de projeto que requerem uma re-inspeção.
  - Conduzirão a mudanças na prática atual da engenharia de software ou trarão uma confiança maior no valor da prática atual. Por exemplo, os pesquisadores e os profissionais gostariam de saber sob que circunstâncias um projeto pode com segurança adotar tecnologias ágeis e sob que circunstâncias não deve.
  - Identificam discrepâncias entre o folclore e a realidade.

### 3.3.1 Definir objetivos de análise

Depois de definir as questões para a análise, o próximo passo é traduzir as questões de interesse em objetivos de análise. Cada questão de interesse pode originar um ou mais objetivos de análise. Os objetivos de análise permitirão que a análise dos dados seja baseada em uma estrutura bem definida. Na SecESE é usado um arcabouço adaptado do GQM para estruturar os objetivos de análise de estudos experimentais, denominado “Objetivo de Análise”. Cada objetivo é descrito em termos de cinco facetas: objeto de análise, propósito, foco, restrições de escopo e ponto de vista. O modelo tem a seguinte forma:

- **Analisar ‘objeto de análise’ para investigar ‘propósito’ com respeito a ‘foco’ no escopo de ‘restrição de contexto’ do ponto de vista de ‘ponto de vista’.**

**Onde:**

- **Objeto de Análise:** define os produtos que serão analisados. Podem ser:
  - Um conjunto de relatórios experimentais;
  - Um conjunto de informações;
  - Um conjunto de dados experimentais.

- **Propósitos** (Tipos de Objetivos): O propósito esboça o que o usuário quer fazer com o objeto do estudo. Prevedemos as seguintes finalidades baseadas no tipo das perguntas genéricas da pesquisa:
  - **Qualidade da Informação**
    1. Caracterizar - avaliar a qualidade da informação ou verificar se a informação coletada atualmente é melhor, similar, ou pior que o esperado quando comparada à informação previamente coletada.
  - **Conteúdo da Informação**
    1. Suporte: confirmar ou rejeitar uma hipótese ou folclore baseado nos resultados apresentados;
    2. Sumarizar: apresentar e resumir os principais resultados encontrados;
    3. Comparar: comparar resultados prévios com resultados atuais.
  - **Extrapolação da Informação**
    1. Explorar - explorar e descobrir padrões previamente desconhecidos na informação analisada;
    2. Predizer - a predição das tendências e dos comportamentos da aplicação dos métodos e técnicas baseados em resultados anteriores.
- **Foco** são as variáveis nas quais o analista do domínio está interessado.
- **Restrições de Contexto:** São o escopo específico de contexto no qual os pesquisadores estão interessados.
- **Ponto de Vista:** Existem múltiplas perspectivas de visualização dos resultados dos estudos experimentais em ES. Em relação às perspectivas de análise, vislumbramos três diferentes perspectivas a serem assumidas durante análise: especialista no domínio dos dados, especialista em experimentação e especialista em análise dos dados. Estas perspectivas são complementares entre si como descritas a seguir:
  - **Especialista em Domínio dos dados** - o especialista no domínio dos dados pode ser definido como um indivíduo que já trabalhou ou trabalha com dados similares aos utilizados na análise, ou tem experiência prática na área de estudo dos estudos

- 
- experimentais. Seu papel é verificar se a análise está sendo realizada por uma interpretação correta sobre a natureza dos resultados.
- **Especialista em Experimentação** - o especialista em experimentação é o indivíduo com reconhecida experiência na área de experimentação e que tem condições de verificar a validade interna e externa dos resultados que estão sendo tratados, assim como de avaliar as questões que podem explicar divergências entre resultados nos diferentes contextos.
  - **Especialista em análise de dados** - é o especialista nas técnicas de análise de dados e que auxiliará na validação das análises e na interpretação correta dos padrões observados nos dados.

Um exemplo de objetivo para extrapolação da informação é:

- **Analisar ‘artigos’ para ‘Explorá-los’ com respeito à ‘Tipos de Defeitos de Interface’ no escopo de ‘estudos experimentais em defeitos’ do ponto de vista de ‘especialistas em engenharia de software, experimentação e análise de dados’.**

É importante observar que apesar de nosso modelo ser baseado no GQM, a semântica é diferente. O modelo proposto por Basili e outros [133] foi especificamente definido para medição em ES e o modelo Na SecESE é voltado à análise de informações.

Na abordagem SecESE, cada estrutura do objetivo especifica os objetivos associados a uma determinada pergunta da análise. Pode-se estabelecer três tipos de objetivos da análise: qualidade da informação, conteúdo da informação e extrapolação da análise da informação. Por esta razão dizemos que estamos estabelecendo uma estrutura de GQF (Objetivos, questões e fatores) ao invés de GQM (Objetivos, Questões e Métricas). Nesta estrutura fornecemos uma plataforma para interpretar a informação, compreender melhor o conhecimento que o usuário necessita, e ajudá-lo a identificar a informação que será extraída.

Desta forma, uma vez que o objetivo foi definido, o analista necessita estabelecer os fatores de interesse nos objetos de análise. Os fatores de interesse são as variáveis que serão efetivamente estudadas. Usamos o modelo mostrado na Tabela 5 para indicar o foco de interesse da análise; este modelo é similar ao usado por Solingen para o GQM [114]:

- **Detalhes do foco:** descreve o foco de análise que o usuário do conhecimento quer fazer. Permite especificar as definições e quaisquer outros detalhes que caracterizem o foco de maneira a facilitar ao analista a coleta dos artigos e fontes de dados que são realmente relevantes ao objetivo da análise.
- **Fatores de variação do foco:** Os fatores que o usuário do conhecimento reconhece como intervenientes sobre o foco do estudo. Se possível deve-se listar a existência de relacionamentos ou efeitos conhecidos destes fatores sobre o foco.
- **Fatores de Interesse:** Dentre os fatores de variação, quais serão investigados na análise.
- **Hipóteses Iniciais:** A opinião e/ou as hipóteses sobre o foco de estudo.

A maneira como o modelo é preenchido também depende do tipo de pergunta da análise definida pelo usuário, a Tabela 6 apresenta um exemplo de detalhamento dos objetivos de pesquisa. A partir das perguntas da análise, o analista sabe quais fatores são os mais importantes para a análise. Estes fatores serão restrições do foco para a análise. O analista pode preencher este modelo de duas maneiras:

- 1) Entrevistando o usuário da análise; ou,
- 2) Procurando pelas informações nos documentos disponíveis sobre o domínio em questão.

**Tabela 5 - Estrutura para Detalhar Objetivos de Análise**

<b><u>Objetivo:</u></b>	
<b><u>Detalhes do Foco</u></b> (Descreve o foco, definições, forma de coleta, etc.).	<b><u>Fatores de Variação</u></b> (Que fatores podem ter impacto no foco?).
<b><u>Fatores de Interesse</u></b> (Dentre os fatores de variação, quais serão investigados na análise).	
<b><u>Hipóteses de Base ou Hipóteses Iniciais</u></b> (O que o usuário acredita ser verdadeiro sobre o foco, relacionado aos fatores de variação).	

Tabela 6 – Exemplo de detalhamento de Objetivos de Análise

<p><b>Objetivo:</b>  <b>Analisar ‘artigos’ para ‘Explorá-los’ com respeito a ‘Tipos de Defeitos de Interface’ no escopo de ‘estudos experimentais em defeitos’ do ponto de vista de ‘especialistas em engenharia de software, experimentação e análise de dados’.</b></p>	
<p><b>Detalhes do Foco</b></p> <p>Definição para Defeitos de Interface: o número de componentes modificados quando uma correção é realizada, isto é, um defeito é um defeito de interface se mais de um componente (ou módulo) deve ser modificado para corrigir um problema.</p>	<p><b>Fatores de Variação</b></p> <ol style="list-style-type: none"> <li>1.Experiência no Desenvolvimento do tipo do sistema;</li> <li>2.Interpretação equivocada dos requisitos;</li> <li>3.Tipo de Projeto.</li> <li>4.Fase de Desenvolvimento</li> </ol>
<p><b>Fatores de Interesse</b></p> <ol style="list-style-type: none"> <li>1. Experiência no desenvolvimento do tipo do sistema.</li> <li>2. Interpretação equivocada dos requisitos;</li> </ol>	
<p><b>Hipóteses Base</b></p> <ol style="list-style-type: none"> <li>2. <b>Interpretação equivocada dos requisitos é a grande causadora dos defeitos de interface.</b></li> <li>3. <b>A grande maioria dos defeitos são defeitos de interface.</b></li> </ol>	

### 3.3.2 Selecionar Artigos

A SecESE usa uma estratégia repetível e eficiente para localizar e selecionar os artigos de interesse baseada nas estratégias usadas nos processos de revisões sistemáticas para realizar esta atividade [11, 12, 16, 30, 74]. O processo é composto das seguintes atividades, descrito nas seguintes subseções:

1. Determine quais bases de dados devem ser usadas na busca;
2. Desenvolva e documente uma estratégia de pesquisa;
3. Conduza a busca.

#### 3.3.2.1 Determine quais bases de dados devem ser usadas na busca.

A busca por estudos relevantes pode ser feita inicialmente nas bases eletrônicas; em Engenharia de Software esta busca deve ser suficiente para selecionar boa parte dos artigos em potencial. Uma busca completa incluirá geralmente buscas em várias fontes de literatura. As fontes da evidência da pesquisa podem ser, por exemplo:

- Bases de referência bibliográficas como IEEExplore etc;
- Listas de Referência a relevantes artigos primários;
- Periódicos, anais de conferências, artigos não publicados;
- Registros de pesquisa;
- Pesquisadores e profissionais na área;
- A internet.

### 3.3.2.2 Desenvolva e documente a estratégia de busca.

Uma vez que foram estabelecidas as bases de dados que serão focadas, necessita-se então definir uma estratégia de busca nestas bases. Na SecESE usa-se a estrutura detalhada dos objetivos de análise para se estabelecer quais são as palavras-chave a serem usadas no processo de busca. O foco, assim como as restrições de contexto podem ser usadas como palavras-chave, sempre lembrando que os sinônimos devem ser considerados. Os detalhes do foco, os fatores de interesse, fatores de variação e hipóteses iniciais também devem ser considerados como fontes de termos para busca. O processo é o seguinte:

Para cada objetivo de pesquisa:

- Selecionar foco e restrições de contexto como possíveis termos de busca;
- Selecionar detalhes do foco, os fatores de interesse, fatores de variação e hipóteses iniciais como possíveis termos de busca;
- Identificar sinônimos, variantes de escrita, e termos mais abrangentes associados a cada um dos termos selecionados nos passos anteriores.

Para ser transparente e replicável é necessário que:

- A revisão seja documentada em detalhe suficiente para que leitores possam avaliar o processo da busca;
- A busca seja documentada à medida que ela ocorre e mudanças devem ser anotadas e justificadas;

- Os resultados não usados da busca da busca devem ser conservados e retidos para uma possível re-análise.

Os procedimentos para documentar o processo de busca são mostrados na Tabela 7 que contém o que deve ser coletado para cada fonte de evidência [16].

**Tabela 7 – Documentação do Processo de Busca adaptado de [16]**

Fonte	Documentação
Bases de Dados Eletrônicas	Nome da Base de Dados Estratégia de Busca Data da Busca Anos cobertos pela busca
Buscas em Periódicos	Nome do Periódico Anos cobertos pela busca Exemplares não incluídos
Anais de Conferências	Título da Conferência Nome da Conferência (se diferente) Nome do periódico (se publicado como parte de um periódico)
Esforços para Identificar materiais não publicados.	Grupos de pesquisa contatados (Nomes e detalhes do contato) Sites de pesquisa pesquisados (Data e URL)
Referências Bibliográficas	Artigo original Artigos escolhidos para Análise
Outras Fontes	Data da Busca URL Quaisquer outros detalhes importantes para a repetibilidade do processo

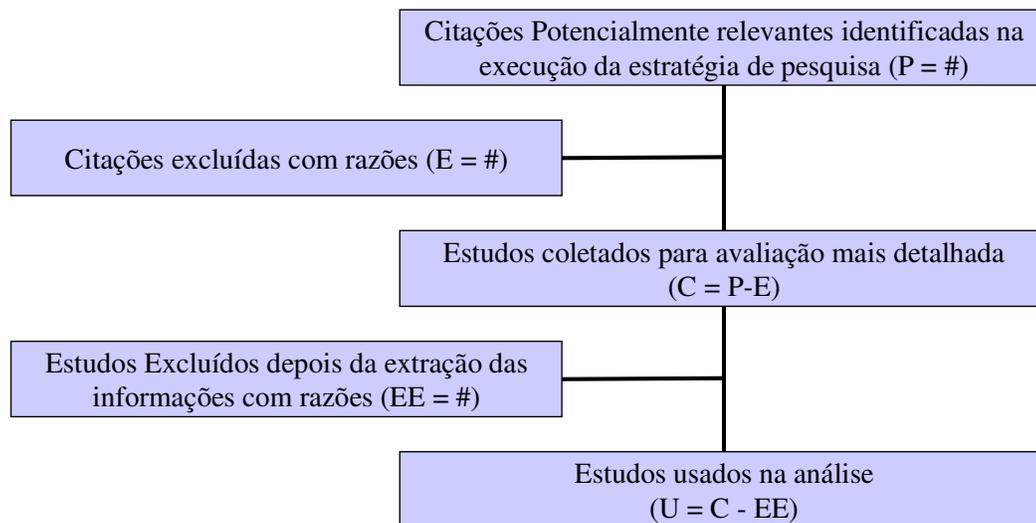
### 3.3.2.3 Conduza a busca.

Uma vez que tem-se uma estratégia de busca e sabe-se onde procurar pelos artigos que serão incluídos na análise, necessita-se definir os critérios de inclusão para determinar que artigos da busca devem ser usados na análise; geralmente pesquisadores fazem o seguinte:

- Executam a estratégia de busca estabelecida no passo anterior em todas as fontes de evidência selecionadas.
- Em cada fonte usam o critério de inclusão para determinar quais artigos devem ser incluídos na análise.

Este processo é iterativo e pode ser visto como mostrado na Figura 12. Na primeira triagem, os pesquisadores usam, geralmente, os critérios de inclusão enquanto analisam os títulos e os resumos dos artigos. Um critério comum de inclusão é avaliar se o conteúdo do artigo está

relacionado ao foco estabelecido no objetivo da análise. Alguns pesquisadores vão um pouco além e fazem um exame na estrutura do artigo para ver se estes estão realmente relacionados à análise, ou se contêm alguns aspectos de interesse do pesquisador, como o método de pesquisa utilizado no artigo. Assim, o número final de artigos a serem analisados vai mudando à medida que novas fontes vão sendo adicionadas e que os critérios de inclusão vão sendo aplicados a estes artigos.



**Figura 12 - Diagrama de Fluxos do Processo de Seleção de Estudos**

### 3.3.3 Definir metodologia da análise

Além de detalhar os objetivos da análise de dados e selecionar os artigos, o analista necessita ainda definir a metodologia para chegar a estes objetivos. O plano de análise deve indicar claramente os seus objetivos e listar as tarefas mais importantes para sua realização. Além destas etapas essenciais, este plano deve também fornecer uma descrição detalhada de exatamente o que o analista quer fazer. Kitchenham e outros e também Biolchini e outros [16, 65] definiram um protocolo específico para a fase de planejamento em uma revisão sistemática.

Na SecESE, o seguinte conjunto mínimo de informações deve ser explicitamente declarado no plano de análise:

- 
- **ID do Analista:** especifica o Analista de dados;
  - **Data da Análise:** especifica uma data de referência para a análise;
  - **Objetivos da Análise:** lista dos objetivos definidos e detalhados na primeira etapa;
  - **Metodologia da Análise:** a metodologia geral usada para análise dos objetos de análise.
  - **Análise de Completude:** impressões do analista sobre a completude da análise em termos de número de estudos sendo analisados. Isto pode incluir: critérios de seleção dos artigos e critérios para inclusão e exclusão dos artigos;
  - **Lista Inicial de Artigos para Análise:** lista de artigos selecionados para avaliação mais detalhada.
  - **Riscos à Análise:** descrição dos potenciais equívocos que o analista já percebe que pode ocorrer na análise.

### 3.4 Avaliação de Qualidade

Uma vez que a informação foi extraída, uma avaliação mais detalhada da sua qualidade necessita ser feita. Esta avaliação é importante para a interpretação dos resultados e auxilia os pesquisadores na avaliação das diferenças nas execuções dos estudos.

A abordagem SecESE para a avaliação da qualidade do estudo é baseada nos conceitos de validade externa e inspirada em uma abordagem usada internamente no Fraunhofer Maryland Center em seu modelo da confiança do Clearing House<sup>1</sup>. Na SecESE cada estudo é avaliado segundo uma série de critérios (Tabela 8). Para cada critério o estudo é avaliado em uma escala numérica. Além dos critérios estabelecidos no centro de Fraunhofer, medimos quão bem detalhadas foram as descrições do contexto dos estudos nos artigos. Os critérios principais são os seguintes:

- 1) *Como a prática foi aplicada:* este critério auxilia na medição da aplicabilidade dos resultados;

---

<sup>1</sup> <http://fc-md.umd.edu/> - Fonte: Forrest Shull – Clearing House Project

- 2) *Como os resultados foram medidos:* este critério auxilia na avaliação do rigor de realização do estudo;
- 3) *Como a experiência foi relatada:* este critério mede a aceitabilidade dos resultados em uma comunidade específica;
- 4) *Pessoa que relatou as evidências:* este critério mede a propriedade da pessoa que relatou os resultados;
- 5) *Grau de detalhe da descrição do contexto:* este critério mede a aplicabilidade da descrição de contexto para a análise.

**Tabela 8 - Critérios de Qualidade**

Como a Prática foi Aplicada?	Não é claro	0
	Um Experimento	3
	Uma série de experimentos	4
	Um projeto piloto	4
	Uma série de projetos pilotos	5
	Um projeto em produção	5
	Uma série de projetos em produção	7
Como os Resultados foram medidos?	Não é claro	0
	Opinião Subjetiva	2
	Estudo de Caso	4
	Comparação entre Projetos	4
	Comparação rigorosa com outra prática ou entre vários projetos	5
Como a experiência foi relatada?	Através de Entrevistas	1
	Não publicado ou somente publicado na Web (white paper).	1
	Opinião de Especialista / Lições Aprendidas	1
	Relatório Técnico (dentro de uma organização ou universidade)	2
	Publicação em Workshop	2
	Publicação em Revista (e.g. CrossTalk)	3
	Publicação em Conferência ou publicação de 2a-ordem (ex. EMSE)	4
Publicação em Periódicos (ex. IEEE TSE)	5	
Pessoa que relatou os resultados	Não é claro	0
	Alguém externo ao time que aplicou a prática	2
	Alguém que foi parte do grupo que aplicou a prática	3
<b>Grau de detalhe da descrição de contexto.</b>	<b>Não descrito ou insuficientes para análise.</b>	<b>0</b>
	<b>Alguns detalhes descritos mas suficiente para análise</b>	<b>2</b>
	<b>Totalmente descrito</b>	<b>4</b>

Para cada critério, existem valores para cada opção estabelecida na Tabela 8; assim, os artigos que de má qualidade podem ser descartados ou considerados de forma diferente no estabelecimento das conclusões. A Figura 13 mostra um exemplo de avaliações de 15 artigos, note-se que alguns artigos, como os artigos 6, 7 e 10, possuem a mesma qualidade parcial, mas o artigo sete não detalha o estudo de forma adequada no artigo, de forma que diminui a usabilidade destes resultados na análise secundária.

Artigo	Como a prática foi aplicada?	Como os resultados foram medidos?	Como a experiência foi reportada?	Pessoa que reportou os resultados	Qualidade Parcial (Max 20)	Grau de detalhe das descrições de contexto	Qualidade Total (Max 25)
1	5	4	4	2	15	4	19
2	5	4	4	2	15	4	19
3	7	4	5	2	18	4	22
4	5	4	4	2	15	2	17
5	5	4	4	2	15	4	19
6	5	4	5	2	16	4	20
7	7	4	5	0	16	0	16
8	5	4	2	0	11	4	15
9	5	0	4	0	9	4	13
10	4	5	5	2	16	4	20
11	7	4	4	0	15	4	19
12	4	4	4	3	15	4	19
13	4	4	4	2	14	4	18
14	7	4	5	2	18	4	22
15	7	0	4	3	14	0	14

**Figura 13 - Exemplo de Avaliação de Qualidade dos Artigos**

Pode acontecer também de alguns critérios serem mais importante que outros para alguns pesquisadores, ou que alguns critérios adicionais, como por exemplo, métodos estatísticos utilizados na análise sejam importante para a análise. Nestes casos o analista pode incluir a priori estes critérios, ou incluir ponderação nos critérios que já estabelecemos aqui. Devem existir dois níveis de qualidade, um global, com critérios globais para qualquer análise que está sendo realizada na base, e um critério específico para a análise, que pode usar uma abordagem similar à usada na revisão sistemática. Exemplos de aplicação destes critérios podem ser encontrados em outras referências [14, 61, 125].

Após a avaliação da qualidade dos artigos, o analista deve reavaliar a lista de artigos a permanecerem na análise. Para isto a analista deve estabelecer um nível mínimo de qualidade para manter o artigo na análise. Artigos que tiverem grau de detalhe da descrição de contexto não descrito ou insuficiente para a análise deve ser fortemente considerado para exclusão.

### 3.5 Extração das Informações

Nesta fase, os objetivos da análise e o seu planejamento já foram realizados, assim como a avaliação da qualidade dos artigos. Têm-se então os relatórios que serão utilizados durante a análise e pode-se então extrair as informações destes objetos de análise. A SecESE usa uma técnica específica para extrair informações de artigos.

A técnica de leitura SecESE tem como objetivo extrair informação codificada (resultados e informações de contexto) dos artigos sob análise. Nesta técnica, o analista segue um procedimento para identificar e coletar resultados e informações de contexto, como segue:

- 1) O artigo é lido mantendo em mente dois tipos de informação que se deseja identificar:
  - a) Resultados; b) Descrições de Contexto.
- 2) Quando as informações relevantes são encontradas durante a leitura, ela deve ser destacada no texto para permitir a rastreabilidade entre as informações coletadas e a sua localização no artigo.
- 3) Transfira as informações requeridas para os formulários de coleta de informações da técnica. A completa descrição dos seus campos está descrita nas subseções seguintes.

#### 3.5.1 Definições

Um **resultado** é uma tentativa de explanação para determinados comportamentos, fenômenos, ou eventos que ocorreram durante um estudo experimental. Um bom resultado indica clara e concisamente o relacionamento entre duas ou mais variáveis e define essas variáveis em termos mensuráveis. Na SecESE, os resultados são divididos em:

- Resultados do Estudo – uma tentativa de explanação para certos comportamentos, fenômenos, ou eventos que ocorreram no estudo experimental relatado.

- 
- Resultados de Outros Estudos Experimentais – um resultado de um estudo que foi relatado em outro artigo e relatado no artigo que está sendo revisado.

As descrições de contexto são os detalhes a respeito do ambiente no qual as medidas foram extraídas. As descrições do contexto são importantes para a compreensão das medidas e dos fatores que influenciam o foco do estudo. A SecESE propõe um formulário padrão para coletar a informação do contexto, que deve ser revisado e adaptado a cada nova análise para verificar se está-se coletando as informações necessárias para a análise dos artigos.

### **3.5.2 Diretrizes Gerais para a Localização das Informações**

Os artigos experimentais são divididos em seções que quase sempre contêm ao menos: introdução, uma descrição do método usado no estudo, uma análise dos resultados, interpretação dos resultados, trabalhos relacionados, e conclusões.

A “introdução” do artigo geralmente posiciona a pesquisa em um contexto, fornecendo uma visão geral e apresentando as principais hipóteses da pesquisa. Estas hipóteses são a base para coleta dos principais resultados relatados no artigo (vide Seção 3.5.3).

A seção do “método” geralmente contém uma descrição de como a pesquisa foi conduzida, incluindo detalhes do projeto experimental, dos participantes, o projeto do estudo, as tarefas dos participantes, e as medidas usadas. Essa seção normalmente contém a maior parte da informação sobre a descrição de contexto requerida pela SecESE (vide definição na Seção 3.5.4). Em cada artigo, deve haver pelo menos uma descrição de contexto associada com cada fonte de informação (talvez mais, se o artigo apresentar informações coletadas de diversos estudos experimentais ou projetos). A seção que contém a “análise dos resultados” geralmente descreve os resultados do estudo (vide definição na Seção 3.5.3). Geralmente, estas seções contêm muitos resultados que não foram anteriormente mencionados no artigo. Algumas vezes uma seção de discussão segue a análise de resultados e contém as interpretações e as implicações do estudo e novos resultados também podem ser extraídos desta seção.

A seção de trabalhos relacionados busca posicionar e comparar o estudo atual com outros estudos executados previamente. Ela geralmente contém os resultados de estudos experimentais relatados em outros artigos (vide definição na Seção 3.5.3).

Na seção de conclusão podem-se encontrar os principais resultados do estudo, embora seja provável que estes resultados já tenham sido relatados em outras partes no texto.

### **3.5.3 Coleta de Resultados**

Os resultados do estudo são usados pelos pesquisadores para responder as questões de pesquisa, e geralmente são encontrados nas seções de análise dos resultados, discussão, trabalhos relacionados ou conclusão. Ao identificar uma frase que o leitor acredita conter um resultado, este deve usar as seguintes perguntas para confirmar se a frase é um resultado:

- Descreve resultados da medição?
- Resume os dados brutos?
- Destaca algumas características específicas dos dados brutos?
- Fornece introspecções sobre tabelas e figuras do artigo?
- Resume os resultados das análises estatísticas?
- Pode ser usado para responder às questões de pesquisa?
- Estabelece respostas para as hipóteses iniciais?
- Reflete os resultados principais do estudo?

Se a resposta for SIM para qualquer das perguntas a frase deve ser coletada. É importante observar que os resultados mais importantes a serem recolhidos são os resultados que podem ser generalizados além do contexto do estudo. Esta avaliação nem sempre é fácil de fazer, pois: (1) nem todos os artigos contêm todas as informações requeridas pelo modelo de contexto da SecESE, e (2) nem sempre é possível inferir se o resultado pode ser generalizado além do contexto no qual ele foi obtido.

Alguns resultados podem ser encontrados também nas tabelas e nas figuras do artigo apesar de não serem explicitamente apontados no texto do artigo. Estes resultados também devem ser inseridos nos resultados deixando claro que foram deduzidos pelo leitor.

Como referência, a Tabela 9 contém alguns números que recolhemos sobre a coleta de resultados em 22 artigos [44]. Note que se o leitor focar a extração de resultados nas seções de Discussão, Conclusão e Análise dos Resultados ele encontrará 89% dos resultados do estudo. No que diz respeito a identificar resultados de outros estudos, se o leitor focar nas seções de introdução e trabalhos relacionados ele deve encontrar 70% dos resultados de outros artigos. Precisamos deixar claro que a Tabela 9 é somente um sumário de um apanhado representativo de artigos da área; os resultados podem variar para artigos específicos.

**Tabela 9 - Resumo da extração de resultados de artigos**

	<b>Resultados Testados</b>	<b>Resultados de Outros Artigos</b>
Total de Resultados Analisados	<b>164</b>	<b>72</b>
Discussão e Conclusão	<b>32.93%</b>	<b>5.56%</b>
Resultados Experimentais	<b>56.10%</b>	<b>15.28%</b>
Introdução		<b>20.83%</b>
Método	<b>9.15%</b>	<b>8.33%</b>
Trabalhos Relacionados		<b>50.00%</b>

### 3.5.3.1 Padrão para Coleta de Resultados

Como diretriz, identificamos as seguintes informações a serem coletadas de cada resultado:

- **Texto:**
  - O resultado deve ser coletado usando o mesmo texto encontrado no artigo de forma a assegurar a sua rastreabilidade em relação à sua posição no texto.
- **Origem:**
  - Nome da seção, figura ou tabela de referência na qual o resultado foi coletado.
- **Tipo:**
  - Dois tipos são possíveis: Resultado do Estudo ou Resultado de Outro Estudo (um resultado que foi testado em outro artigo e relatado no artigo que está sendo lido).
- **Nível de Suporte:**
  - O avaliador deve explicitamente indicar o nível de suporte apresentado no artigo para o resultado coletado. O suporte deve ser descrito como um dos seguintes níveis:

- **Significativamente Positivo:** Os resultados são estatisticamente significativos, isto é, com um alto grau de certeza e não são resultantes de pura chance.
  - **Positivo:** Os resultados são estatisticamente significativos, mas sem grande grau de liberdade estatística.
  - **Opinião:** Uma frase explicativa para certos comportamentos, fenômenos, ou eventos que são formulados pelos autores do artigo. Apesar de estar no formato de um resultado, esta frase é fundamentalmente baseada em suposições ou conhecimentos prévios dos autores, e estes não apresentam evidências experimentais que suportem esta afirmação.
- **Observações:**
    - **Texto livre.** O leitor pode incluir algumas informações adicionais sobre o resultado. Estas informações podem ser observações que os autores fizeram no texto e também opiniões próprias do leitor sobre o resultado, que podem auxiliar posteriormente no processo de análise deste resultado.

### **3.5.4 Coleta de Informações de Contexto.**

A SecESE define um modelo sistemático para coletar a informação de contexto dos artigos baseado no artigo de Sjoberg e outros [32]. As descrições do contexto são os detalhes a respeito do ambiente no qual as medidas e os resultados foram obtidos. Estes detalhes são coletados a partir de um formulário padrão (template) extensível que lista as principais facetas que podem ser usadas para caracterizar um estudo experimental em engenharia de software. Este formulário deve ser estendido se necessário antes do início da extração das informações dos artigos.

#### **3.5.4.1 Formulário Padrão para Coleta de Informações de Contexto**

A Figura 14 mostra o formulário padrão de coleta de informações de contexto. Haverá pelo menos um formulário preenchido para cada artigo (talvez mais, se o artigo descrever os dados coletados de diversos estudos ou projetos). Nem todos os campos são obrigatórios, mas são desejáveis. Os valores faltantes necessitarão ser justificados durante a análise e limitam a força das conclusões que podem ser extraídas.

<b>Título do Artigo</b>		
<b>Tópico</b>		
<b>Tipo de Estudo</b>		
<b>Objetivos</b>	Analisar	
	Com o propósito de	
	com respeito à	
	do ponto de vista de	
<b>Variáveis</b>	Nome:	
	Tipo:	
	Valores Possíveis	
	Procedimentos de Coleta dos Dados	
	Nome:	
	Tipo:	
	Valores Possíveis	
	Procedimentos de Coleta dos Dados	
<b>Participantes</b>	Categoria:	
	Número:	
<b>Instrumentação</b>	Nome da Ferramenta:	
	Funções:	
	Nome da Ferramenta:	
	Funções:	
<b>Tarefas</b>	Categoria:	
	Duração	
	Modo de Trabalho:	
<b>Produtos de Trabalho</b>	Nome:	
	Tipo:	
	Origem:	
	Domínio da Aplicação:	
	Tamanho:	
	Paradigma de Representação:	
Linguagem:		
<b>Replicação</b>		
<b>Outros</b>		

Figura 14 - Formulário de Descrição de Contexto.

Os atributos do formulário de descrição de contexto são:

- **Título do Artigo:**

- O título do artigo do qual o leitor está extraindo a informação.

- **Tópico:**

- Para derivar as categorias do tópico temos usado as palavras chaves da Sociedade de Computação do IEEE disponíveis no website Computer.org, deve-se observar que esta lista pode não conter os tópicos de interesse para o analista, e, portanto, deve ser revista e adaptada para as necessidades da análise. Os principais tópicos são os seguintes (esta lista pode ser estendida usando a lista completa disponível no site da IEEE<sup>2</sup>):

- Engenharia de Software – Geral;
- Requisitos/Especificação;
- Ferramentas e Técnicas de Projeto;
- Ferramentas e Técnicas de Codificação;
- Verificação de Programas;
- Teste e Depuração;
- Ambientes de Programação e Construção;
- Distribuição, Manutenção e Melhorias;
- Métricas e Medição;
- Gerenciamento;
- Projeto;
- Arquitetura de Software;
- Interoperabilidade;
- Reuso de Software;
- Fatores Humanos em Projetos de Software;
- Segurança de Sistemas e Software;
- Gerencia de Configuração;
- Construção de Software;
- Processos de Engenharia de Software;
- Qualidade de Software/ SQA.

- **Tipo de Estudo**

- **Experimento** - Uma investigação detalhada e formal que é executada sob circunstâncias controladas. O objetivo é manipular uma ou mais variáveis e

---

<sup>2</sup> <http://www.computer.org/mc/keywords/software.htm>

---

controlar as variáveis restantes. A finalidade de um experimento controlado é fazer as observações cujas causas são não-ambíguas. Isto é alcançado isolando-se os efeitos de cada fator dos efeitos de outros fatores, podendo então estabelecer-se corretamente as relações de causa e efeito [29].

- **Estudo de Caso** - Uma investigação detalhada de um único “caso” ou de “casos relacionados”. Tal investigação é executada em condições típicas; por exemplo, a partir de alguns projetos típicos representativos. Um estudo de caso é um estudo observacional enquanto um experimento é um estudo controlado [94].
- **Pesquisa de Opinião** - Uma investigação mais ampla onde a informação é coletada em um formulário padrão preenchido por um grupo de pessoas ou de projetos após o fato. Os meios principais de se recolher dados qualitativos ou quantitativos são entrevistas e questionários.

• **Objetivos:**

- Descreve os objetivos para o estudo descrito no artigo usando o padrão GQM [133]. Este padrão (template) tem o seguinte formato:
  - **Analisar objeto de estudo**
    - Objeto de estudo: principal entidade de interesse do estudo, isto é, o processo, produto, modelo, ou objeto que é estudado.
  - **Com o propósito de X**
    - X: finalidade do estudo em relação ao seu objeto de análise, isto é, o estudo objetiva caracterizar, entender, avaliar, predizer, ou aprimorar o objeto de estudo.
  - **Com respeito a M**
    - M: atributos de interesse do objeto de análise neste estudo, isto é, sob que aspecto pretende-se analisar o objeto de estudo (efetividade, número de defeitos, tempo de desenvolvimento, etc.).
  - **Do ponto de vista de P**
    - P: para quem o estudo deve ser de valor, isto é, o tipo de pesquisador, gerente de projetos, ou profissional que tem interesse no estudo.

- Exemplo de Objetivo:

<b>Objetivos</b>	<i>Analisar</i>	Técnicas de leitura de código, teste funcional e teste estrutural.
	<i>Com o propósito de</i>	Avaliá-las
	<i>Com respeito a</i>	# falhas observadas & tempo por defeito
	<i>Do ponto de vista do</i>	Pesquisador

- **Variáveis**

- Uma variável é um conceito ou uma construção que pode variar ou ter mais de um valor. O pesquisador pode estar interessado em saber como determinadas variáveis estão relacionadas. Por exemplo, que variáveis afetam a “eficácia” de uma técnica de teste? Ou, pode estar interessado em compreender o relacionamento entre o número dos defeitos e o tamanho dos programas.
- Existem dois tipos básicos de variável.
  - A variável independente é definida como a “variável que o experimentador manipula”. Isto é válido para experimentos, mas nem todos os estudos são experimentos. Em alguns tipos de estudos, os pesquisadores não manipulam uma variável em um estudo. Em vez disto, somente coletam dados e observam como as variáveis estão relacionadas entre si. A variável independente nestes casos é aspecto que o investigador está estudando com respeito a como se relaciona a ou influencia outras variáveis (as variáveis dependentes). Se a variável independente estiver relacionada ou influenciar a variável dependente, ela pode ser usada para prever a variável dependente. Conseqüentemente, uma variável dependente é chamada às vezes de variável preditora ou variável explanatória. A variável independente pode ser manipulada ou pode apenas ser medida.
  - A variável dependente é o que o pesquisador está estudando, com respeito a como se relaciona ou é influenciado pela variável independente ou a como pode ser explicada ou estimada pela variável independente. É chamada às

---

vezes de variável resposta. Esta variável nunca é manipulada como uma parte do estudo.

- Uma sugestão útil para determinar o tipo da variável em um estudo é perguntar se está se tentando influenciar ou predizer uma variável a partir de alguma outra variável ou outras variáveis. Quando se está tentando predizer a variável, ela é provavelmente uma variável dependente. A variável que o pesquisador está usando para fazer as predições ou para determinar se ela influencia outra variável no estudo, é uma variável independente.
- As seguintes características são coletadas para cada variável:
  - *Nome*: nome pelo qual a variável é referenciada no artigo.
  - *Tipo*: tipo da variável: independente, dependente, ambos (dependente e independente);
  - *Valores Possíveis*: Domínio de valores possíveis para a variável, se especificado.
  - *Detalhes do Processo de Coleta*: detalhes do método usado para medir a variável, incluindo, por exemplo, que instrumentação e ferramentas de apoio foram usadas.

• **Participantes.**

- São os executores do estudo:
  - Se o experimento é totalmente automático, uma ferramenta deve ser o executor da tarefa e provavelmente será descrito na seção de instrumentação. Neste caso este campo deve ser deixado em branco.
  - Se o experimento não for totalmente automático, um participante deve ser o executor da tarefa:
    - *Categoria*: Uma descrição generalizada do nível da experiência dos participantes. Valores possíveis:
      - Estudantes de Graduação;
      - Estudantes de Pós Graduação;
      - Estudantes (tipo não caracterizado);

- Profissionais;
- Cientistas;
- Outros;
- Desconhecido (tipo não descrito no artigo)
- Número: Número de participantes no estudo experimental.

- **Instrumentação**

- Ferramentas automáticas de medição ou ferramentas auxiliares que geram dados para o estudo. Ou seja, o ferramental que é usado na tarefa de gerar dados.
  - *Nome*: nome do programa;
  - *Descrição das Funções*: descrição das funções realizadas pela ferramenta no estudo experimental.

- **Tarefas**

- *Categoria*: Descreve as tarefas dadas para os “executores” de acordo com as **tarefas** aplicadas e os **produtos de trabalho** nos quais eles foram aplicados (ex. criar um documento de projeto).
  - Possíveis valores das **tarefas**
    - Planejar;
    - Criar;
    - Modificar;
    - Analisar.
  - Possíveis **produtos de trabalho**
    - Requisitos;
    - Projeto/Arquitetura;
    - Código;
    - Relatórios de Mudanças;
    - Relatórios de Defeitos;
    - Etc.
- *Duração*: tempo de duração da tarefa.

- 
- *Modo de Trabalho*: especifica se os participantes executaram as tarefas:
    - Em time;
    - Individualmente.

- **Produtos de Trabalho.**

- Descrição dos artefatos usados nas tarefas. Eles podem ser, por exemplo, documento de especificação ou código. Geralmente cada estudo experimental usa mais de um produto de trabalho.
- *Nome*: estabelece um nome para o produto, geralmente o nome com o qual ele é referenciado no texto.
- *Tipo*: tipos possíveis de produtos de trabalho:
  - Requisitos;
  - Projeto/Arquitetura;
  - Código;
  - Relatórios de Mudanças;
  - Relatórios de Defeitos;
  - Etc.
- *Origem*: descreve a origem do produto de trabalho usado nas tarefas. Os valores possíveis são:
  - Construído: artefato construído para o propósito do estudo experimental;
  - Comercial: um artefato comercial;
  - Projeto de Estudante: um artefato construído para um trabalho de classe;
  - Código Aberto: um artefato de domínio público (ex., aplicação de código aberto);
  - Outros;
  - Não está claro.
- *Domínio do Artefato*:
  - Processamento de Texto
  - Simulação de Vôo

- Etc...
- *Tamanho:*
  - Tamanho do artefato, usando a métrica especificada pelo autor. Ex: 129 linhas de código, de 2000 linhas executáveis de código, 31 páginas, etc.
- *Paradigma de Representação*
  - Os paradigmas de representação dos produtos de trabalho. Exemplo: Orientado a Objetos, Imperativo, Declarativo, Linguagem Natural.
- *Linguagem*
  - Linguagem usada para escrever o produto de trabalho. Exemplo: Inglês, Fortran, Pascal, C++, Java.
- **Repetição:** Indicativo de se o estudo é uma repetição de outro estudo. Os valores possíveis são “Sim” e “Não”. Se este for uma repetição, deve-se incluir a referência ao estudo experimental original e as diferenças entre as replicações.
- **Outros:** Qualquer outra informação que seja importante para o entendimento do contexto do estudo experimental (ex: definições implícitas, características de ambiente, informações sobre conformidade do processo, informações sobre a análise feita nos estudos).

Este formulário padrão deve ser revisado e/ou expandido cada vez que se desejar iniciar uma análise, para certificar-se de que todos os aspectos importantes do estudo estão sendo coletados.

### 3.6 Análise de Resultados

A entrada para esta atividade é uma base de informações (que é um subconjunto da base de conhecimento) sobre os estudos. Esta base contém os resultados, as definições e as descrições do contexto dos estudos assim como a avaliação de sua qualidade. A saída da atividade depende dos objetivos do investigador para a análise. Os resultados podem ser:

- Hipóteses novas, sem evidências que as suportam, mas que podem ser investigadas com novos estudos experimentais;

- 
- Confirmações de Hipóteses;
  - Generalizações de contexto para hipóteses iniciais;
  - Limitações de contexto para algum resultado específico;
  - Validação de Folclores ou de um conjunto inicial de hipóteses.

Esta atividade é composta das seguintes etapas: Formalização dos Resultados e Interpretação dos Resultados.

### **3.6.1 Formalização dos Resultados**

Na etapa da formalização de resultados, as informações coletadas nos artigos são formalizadas de maneira a viabilizar e facilitar a análise dos resultados experimentais. Para tal, definimos resultados como sendo relacionamentos entre duas variáveis (uma dependente e uma independente). Nos casos onde os resultados, como indicados no artigo fonte, descrevem múltiplos relacionamentos, os resultados formalizados têm que ser divididos em diversos menores cada um estabelecendo um relacionamento entre duas variáveis.

A formalização dos resultados é uma etapa que deve ser realizada para facilitar a análise dos resultados por meio de ferramentas de mineração de dados. É importante que o pesquisador, ao executar esta etapa, mantenha o foco do estudo, porque este dirigirá a escolha das variáveis dependentes e independentes a serem estudadas.

Existem dois tipos principais de resultados: uma sentença sobre um relacionamento lógico entre duas variáveis (por exemplo, “o tamanho do módulo aumenta, o número dos defeitos introduzidos aumenta”) e uma indicação de uma função matemática que relacione duas variáveis (por exemplo, “defeitos de interface são aproximadamente 25% de todos os defeitos em um sistema”).

Uma versão formalizada de um resultado tem três componentes principais: uma variável dependente, uma variável independente, e uma relação entre eles. As variáveis podem ser qualquer atributo discutido no resultado (por exemplo, “o tamanho do módulo”, “número de defeitos introduzidos”, “número de defeitos de interface”).

Cada uma das variáveis dependentes e independentes tem também três campos adicionais que são opcionais: a direção do efeito, a magnitude do efeito, e as restrições no efeito. Estes campos terão valores somente se o texto coletado contiver a informação. Nos casos onde o resultado descreve um relacionamento lógico, a direção e a magnitude para cada variável capturam informações sobre os efeitos relacionados às mudanças no valor. Por exemplo, se o resultado fosse “enquanto o tamanho do módulo aumenta de 10%, o número de defeitos injetados aumenta em 25%” então a variável independente do “tamanho do módulo” tem a direção “aumento” e a magnitude “10%”. Nos casos onde o resultado descreve uma função matemática, a direção e a magnitude capturam o cálculo funcional. Por exemplo, se o resultado for “25% de todos os defeitos do sistema são defeitos de interface” então a variável dependente é “defeitos de interface” com a magnitude de 25%, mas sem direção.

Uma restrição em qualquer variável captura simplesmente a informação sobre a escala da aplicabilidade da variável. Para o resultado “à medida que o tamanho dos módulos em C aumentam em 10%...”, o campo de restrição para a variável independente conteria a informação que este resultado só é válido para a linguagem de programação de C.

O campo da relação descreve sempre o relacionamento hipotético entre as variáveis independentes e dependentes. Os valores podem ser:  $\approx$  (aproximadamente igual),  $<$ ,  $>$ ,  $\leq$ ,  $\geq$ ,  $=$ ,  $\neq$ ,  $\Rightarrow$  (implica),  $\nrightarrow$  (não implica).

Como ilustração, considere os seguintes exemplos:

1. Texto: Módulos maiores são mais complexos que módulos menores.

a. Formalização:

**i. Tamanho maior nos módulos implica em maior complexidade nos módulos.**

1. Direção Independente: maiores
2. Magnitude da Diferença - Independente: N/A
3. Variável Independente: tamanhos
4. Contexto Independente: módulos

- 
5. Efeito: implicam
  6. Direção Dependente: maior
  7. Magnitude da Diferença - Dependente: N/A
  8. Variável Dependente: complexidade
  9. Contexto Dependente: módulos
- b. Fonte: [Basili84].
  - c. Nível de Suporte: Significativamente Positivo.
  - d. Confiança: Alta.
2. Texto: 89% dos defeitos podem ser corrigidos mudando-se somente um modulo
- a. Formalização:
    - i. **89% das correções de defeitos são iguais a mudanças em um modulo.**
      1. Direção Independente: N/A
      2. Magnitude da Diferença - Independente: 89%
      3. Variável Independente: Correções em Defeitos
      4. Contexto Independente: Módulo
      5. Efeito: Igual
      6. Direção Dependente: N/A
      7. Magnitude da Diferença - Dependente: N/A
      8. Variável Dependente: Mudanças
      9. Contexto Dependente: em um módulo
  - b. Fonte: [Basili84].
  - c. Nível de Suporte: Significativamente Positivo.
  - d. Confiança: Alta.
3. Texto: Existe uma maior incidência de defeitos em módulos de tamanho menores.
- a. Formalização:
    - i. **Menor tamanho dos módulos implica em maior incidência de defeitos.**
      1. Direção Independente: Menor
      2. Magnitude da Diferença - Independente: N/A
      3. Variável Independente: Tamanho

4. Contexto Independente: módulos
  5. Efeito: implica
  6. Direção Dependente: Maior
  7. Magnitude da Diferença - Dependente: N/A
  8. Variável Dependente: defeitos
  9. Contexto Dependente: em número
- b. Fonte: [Basili84]
  - c. Nível de Suporte: Significativamente Positivo
  - d. Confiança: Baixo
  - e. OBS: O baixo grau de confiança vem da maneira na qual os autores explicam os resultados: "A explicação mais plausível parece ser a de que um grande número de defeitos de interface está espalhado igualmente em todos os módulos, causando um grande número de defeitos por 1000 comandos executáveis para módulos pequenos. Outras tentativas de explicação são: a maioria dos módulos examinada eram pequenos, causando um resultado equivocado; módulos maiores foram codificados com mais cuidados do que módulos menores, por causa do tamanho; e, defeitos em módulos menores são mais aparentes" [Basili84].
4. Texto: Erros de entendimento de requisitos e de interface são mais difíceis de corrigir do que outros tipos de erros.
    - a. Formalização:
      - i. Erros de Entendimento de Requisitos implicam aumento na dificuldade de correção do defeito**
        1. Direção Independente: N/A
        2. Magnitude da Diferença - Independente: N/A
        3. Variável Independente: Entendimento de Requisitos
        4. Contexto Independente: Erros
        5. Efeito: Implicam
        6. Direção Dependente: Aumento
        7. Magnitude da Diferença - Dependente: Dificuldade

---

8. Variável Dependente: Correção de Defeitos

9. Contexto Dependente: n/a.

**ii. Erros de Interface implicam aumento na dificuldade de correção do defeito**

1. Direção Independente: N/A

2. Magnitude da Diferença - Independente: N/A

3. Variável Independente: Interface

4. Contexto Independente: Erros

5. Efeito: Implicam

6. Direção Dependente: Aumento

7. Magnitude da Diferença - Dependente: Dificuldade

8. Variável Dependente: Correção de Defeitos

9. Contexto Dependente: n/a

b. Fonte: [Weiss85]

c. Nível de Suporte: Significativamente Positivo

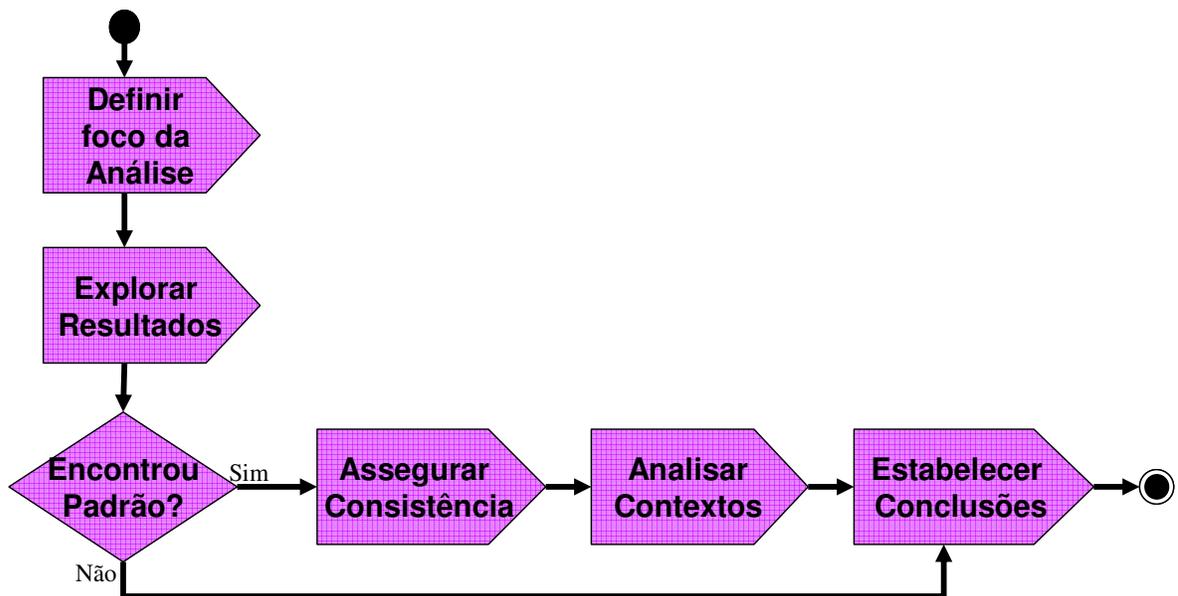
d. Confiança: Alto

Depois da formalização dos resultados tem-se então uma base estruturada de resultados que pode ser usada no processo de análise. A formalização dos resultados é uma atividade de apoio para a atividade de interpretação dos resultados experimentais. Ela transforma todas as informações extraídas para o mesmo padrão de representação, facilitando o manuseio e análise de grandes conjuntos de resultados experimentais. A atividade de formalização é também necessária para viabilizar a automação de algumas atividades que dão apoio à atividade de interpretação.

### **3.6.2 Interpretação dos Resultados**

Nesta fase do processo o analista deve analisar todo o conjunto de informações que foram coletadas para que os objetivos e as questões levantadas no início do processo sejam respondidos. As informações disponíveis são os objetivos da análise, a base de resultados, a caracterização do contexto dos estudos e a avaliação da qualidade dos estudos.

A SecESE propõe um processo iterativo, dentro da atividade de análise de resultados para a subatividade de interpretação de resultados. O processo foca em uma variável dependente por vez; é possível se estabelecer tanto uma análise top-down quanto bottom-up. A análise top-down é feita principalmente quando têm-se hipóteses iniciais que o analista quer avaliar, e a bottom-up quando o analista quer explorar a base de resultados ou construir uma hipótese detalhada. O processo proposto é mostrado na Figura 15.



**Figura 15 - Processo para Interpretação dos Resultados**

As atividades do processo estão descritas na Tabela 10 e serão detalhadas nas subseções que se seguem:

**Tabela 10 - Atividades do Processo para Interpretação dos Resultados**

<b>Atividade: Definir Foco da Análise</b>	
Objetivo	Definir o foco da análise. Esta atividade deve ser feita baseando-se nos objetivos definidos no planejamento da análise (Seção 3.3), no início de todo o processo.
Subatividades	Escolher objetivo da Análise. Escolher variável dependente a ser explorada. Filtrar Resultados baseado no objetivo escolhido.
Produtos	Objetivo de Análise Escolhido Subconjunto de resultados que serão foco da análise.

Responsáveis	Especialista no Domínio
<b>Atividade: Explorar Resultados</b>	
Objetivo	Encontrar padrões nos resultados que possam responder ao objetivo de análise escolhido na atividade anterior. Esta análise é baseada nos fatores de variação, nos fatores de interesse e nas hipóteses de base definidas no planejamento da análise. Estas definições serão importantes para direcionar a exploração dos resultados.
Subatividades	
Produtos	Conjunto de padrões encontrados que podem contemplar os objetivos
Responsáveis	Especialista no Domínio, Especialista em Experimentação e Especialista em Análise
<b>Atividade: Assegurar Consistência</b>	
Objetivo	Assegurar que os padrões encontrados nos resultados são consistentes conceitualmente e semanticamente. Esta etapa é importante pois diminui a probabilidade de se chegar a conclusões equivocadas na próxima etapa.
Subatividades	
Produtos	Padrões nos Resultados
Responsáveis	Especialista no Domínio e Especialista em Experimentação
<b>Atividade: Analisar Contextos</b>	
Objetivo	O analista deve estar ciente dos contextos em que cada um dos resultados nos padrões foram gerados para fazer a interpretação correta e concluir de forma consciente sobre o contexto no qual uma conclusão será incluída. O objetivo aqui é agrupar os contextos dos padrões e delimitar os escopos nos quais os resultados estão inseridos, definindo os limites entre eles. Baseado no contexto de cada resultado, o analista pode verificar a “distância”, a proximidade, a afinidade e a confiança no contexto em que estes resultados foram obtidos e extrair as conclusões baseadas nos padrões observados, criar hipóteses novas, refutar ou confirmar uma hipótese ou folclore inicial. O produto final sempre dependerá do objetivo definido no plano de análise.
Subatividades	Agrupar Contextos
Produtos	Categorização Situacional dos Contextos (iguais, disjuntos, interseccionais).
Responsáveis	Especialistas no Domínio, em Experimentação e em Análise.
<b>Atividade: Estabelecer conclusões</b>	
Objetivo	O objetivo desta atividade é avaliar as conclusões estabelecidas na atividade anterior verificando o interessantismo destas conclusões.
Subatividades	Estabelecer Conclusões
Produtos	Conclusões do Processo de Análise
Responsáveis	Especialistas no Domínio, em Experimentação e em Análise.

### 3.6.2.1 Definir Foco da Análise

O objetivo da atividade de Analisar Resultados (Figura 11) é extrair conclusões sobre o conjunto de informações coletadas e formalizadas durante o processo. Durante a atividade de extração, muitos resultados são coletados, nem sempre relacionados a todos os objetivos de análise. A primeira atividade é portanto definir o foco da análise. Esta atividade deve ser feita baseando-se nos objetivos definidos no planejamento da análise, no início de todo o processo. As subatividades são:

1. Escolher um objetivo de análise a partir do conjunto de objetivos da análise do plano da análise.
2. Filtrar Resultados. Esta filtragem visa diminuir o escopo dos resultados de acordo com o objetivo escolhido. Como a filtragem pode introduzir equívocos, recomendamos que ela seja feita somente de acordo com critérios objetivos. Qualquer filtragem deve ser relatada claramente como uma decisão de análise (nas anotações) para garantir a repetibilidade. A filtragem pode ser feita com relação aos detalhes de contexto ou detalhes dos resultados, ou avaliação da qualidade:
  - Contexto: a filtragem é feita baseada no contexto, o analista pode usar alguns dos campos do modelo do contexto;
  - Resultados: o usuário pode filtrar os resultados baseado em alguns dos campos da formalização do resultado: variáveis dependentes, variáveis independentes etc.
  - Qualidade: o usuário pode escolher por exemplo que o tipo da publicação será considerado, ou estabelecer uma medida parcial ou total mínima de qualidade para incluir um resultado na análise.
3. Escolher uma variável dependente para explorar. Esta escolha é baseada no foco da análise do objetivo escolhido.

### 3.6.2.2 Explorar Resultados

O número de resultados coletados na fase de extração pode ser grande a depender do número de artigos incluídos na análise. Por isto, torna-se necessário o uso de ferramentas que

---

auxiliem o analista na tarefa de exploração dos resultados na busca de respostas às questões de análise. A visualização de informações é, portanto, uma estratégia interessante para ser usada aqui.

O uso de técnicas de visualização de informações tem basicamente três objetivos: análise exploratória, confirmatória e apresentação. Estes objetivos estão relacionados aos propósitos dos objetivos de análise estabelecidos no planejamento da análise, como mostrado na Tabela 11.

**Tabela 11 - Objetivos da Visualização x Propósitos de Análise**

Objetivo da Visualização	Propósito do Objetivo
<b>Análise Exploratória</b>	<b>Sumarizar, Explorar.</b>
<b>Análise Confirmatória</b>	<b>Suporte, Comparar.</b>
<b>Apresentação</b>	<b>Sumarizar</b>

No contexto da SecESE, utiliza-se uma técnica hierárquica de exploração visual de informações para explorar a base de resultados previamente construída. As características dos resultados podem ser associadas a atributos visuais na ferramenta de visualização, seus valores podem ser usados para organizar os diversos resultados na telas, e seus controles interativos permitem selecionar e filtrar os resultados de interesse. Isto cria um ambiente de exploração interativo e dinâmico muito adequado à tarefa de investigação de grandes conjuntos de evidências experimentais.

Estas técnicas permitem que se organizem os dados hierarquicamente, segundo um determinado critério. Na SecESE, isto equivale a organizar as informações de acordo com atributos dinamicamente escolhidos pelo cientista que está explorando a base de resultados. A característica dinâmica deste tipo de ferramenta permite que um perito possa rapidamente chavear contextos de exploração pela modificação dos critérios hierárquicos usados para organizar os resultados sendo explorados.

Como diretriz básica, definimos os seguintes passos para exploração dos resultados:

- 1) Organize os resultados agrupando-os com relação às variáveis dependentes;

- 2) No segundo nível da hierarquia, agrupe cada variável dependente com relação às variáveis independentes que influenciam a variável dependente do passo 1 .
- 3) Se necessário, use os outros campos da formalização dos resultados: Direção, Magnitude, Contexto, e Efeito para agrupar os resultados e auxiliar na exploração de padrões.
- 4) Inclua variáveis do contexto na hierarquia; isto pode ajudar a visualizar alguns detalhes dos resultados;
- 5) Inclua detalhes sobre o critério de qualidade; isto ajuda na avaliação dos resultados que estão sendo combinados;
- 6) Faça várias iterações entre estes campos, tendo sempre a participação do Especialista do Domínio na identificação dos padrões.
- 7) Interaja com estas variáveis, e tente encontrar um padrão nos resultados, ou uma resposta para a questão de análise, ou verificar se os resultados podem suportar ou rejeitar as hipóteses base definidas no objetivo de análise. Esta etapa só é finalizada quando se esgotam todas as possibilidades de análise para aquele objetivo. Se um padrão não é encontrado, o analista vai diretamente para a etapa de Estabelecer Conclusões.

No Capítulo 5 apresentamos um exemplo detalhado de como aplicamos a SecESE na área de defeitos de software, na qual usamos Mapas em Árvore [18, 20] para fazer a exploração dos resultados.

### 3.6.2.3 Assegurar Consistência

Se um padrão é encontrado nos resultados, o pesquisador, baseado em descrições do contexto, deve se certificar de que todos os problemas que podem causar equívocos na análise estão sendo considerados e que os resultados estão sendo corretamente interpretados. Primeiramente o analista deve considerar as “inconsistências” que podem ocorrer entre os

---

estudos. Baseados no trabalho de Visser e outros [108], os tipos de “inconsistências” que prevemos são:

- **Inconsistências Conceituais:** Uma diferença na maneira na qual um domínio é interpretado (conceitualmente), que resulta em conceitos diferentes ou em relações diferentes entre aqueles conceitos. Há dois tipos:
  - 1) **Escopo:** Neste tipo de inconsistência, dois resultados parecem representar o mesmo conceito, mas não têm exatamente o mesmo significado (embora possa haver alguma sobreposição). Por exemplo: Dois estudos podem referenciar o “custo” de uma prática, embora um possa incluir no cálculo somente o custo do esforço para aplicar a prática, o outro pode incluir também os custos de inicialização do processo (como o custo de envio de pessoal para o treinamento).
  - 2) **Cobertura e granulosidade do modelo:** Este tipo de inconsistência descreve os problemas que podem ocorrer em tentar combinar resultados quando não é claro a que parte do domínio aqueles resultados são aplicáveis. Por exemplo, um estudo pode fazer afirmações sobre uma classe grande de projetos de desenvolvimento do software, mas somente tendo a evidência a respeito de um ou dois exemplos específicos de tais projetos.
- **Inconsistências de Explicação:** Diferenças na maneira pela qual a conceitualização é especificada. Estas diferenças podem se manifestar nas inconsistências de definições, nos termos e nas combinações de ambos. Há três tipos:
  - 1) **Termos Sinônimos:** Sinônimos neste contexto são termos diferentes que referenciam o mesmo conceito. Um exemplo trivial é o uso do termo “força” em um estudo e do termo “coesão” em outro, para referenciar ao mesmo conceito (neste caso, a quantidade da interação dentro dos componentes de um sistema). Embora a solução técnica para este tipo de problema pareça relativamente simples (o uso de taxonomias e glossários), a integração dos estudos com sinônimos ou linguagens

diferentes requer geralmente um esforço humano significativo para identificar e resolver as questões semânticas.

- 2) **Termos Homônimos:** Este tipo de inconsistência ocorre quando o significado de um termo é diferente em contextos diferentes. Por exemplo, o termo “defeitos de interface” pode ter interpretações diferentes dependendo do contexto: Pode referenciar um defeito na Interface Humano-Computador ou um defeito nas interfaces entre dois componentes de software. Esta inconsistência é muito mais difícil de lidar, pois conhecimento de diferentes domínios é geralmente requerido para resolver estas ambigüidades.
- 3) **Codificação:** Os valores nos estudos podem ser codificados em formatos diferentes. Por exemplo, os números de linhas do código podem ser representados como “KLOC” ou como “LOC” ou “SLOC”, etc. Se as diferenças estão no fato de que os resultados estão relatados apenas usando níveis diferentes de granulosidade, uma etapa de transformação pode ser aplicada para eliminar todas as diferenças. Se este não for o caso e o analista não conseguir estabelecer os resultados no mesmo nível de granulosidade, ele deve anotar esta questão nas anotações e para que este caso seja lembrado quando estiver estabelecendo as conclusões da análise.

A abordagem da SecESE para a solução de problemas de inconsistências conceituais entres os estudos é simplificada. A consistência entre investigadores diferentes para definir uma estrutura conceitual durante uma revisão de artigos e, conseqüentemente, garantindo a obtenção de resultados comparáveis pode ser aprimorada tendo o apoio de uma terminologia comum e formalizada dos conceitos envolvidos, representada por uma ontologia [66]. Embora as ontologias possam trazer algumas limitações e complexidade ao processo, acreditamos que a SecESE poderia se beneficiar com o uso de ontologias, devido à sua aplicabilidade na formalização e na representação conceitual do conhecimento.

#### 3.6.2.4 Analisar Contextos

Uma vez que padrões são encontrados nos resultados, o analista deve estar ciente dos contextos em que cada um dos resultados foram obtidos para fazer a interpretação correta e

---

consciente sobre o contexto no qual uma conclusão será incluída. Baseado no contexto de cada resultado, o analista pode verificar a “distância”, a proximidade, a afinidade e a confiança no contexto em que estes resultados foram obtidos.

Como discutido na Seção 3.5.4, um contexto é composto de vários atributos. Desta forma, para compará-los é necessário medir a distância entre eles. Mais que isto, em uma situação de vários estudos, caso típico de aplicação da SecESE, o que se deseja é organizá-los em um conjunto de cenários, de forma que cada cenário agrupe um conjunto de estudos com contextos similares. Este é um problema típico de aglomeração estatística [8].

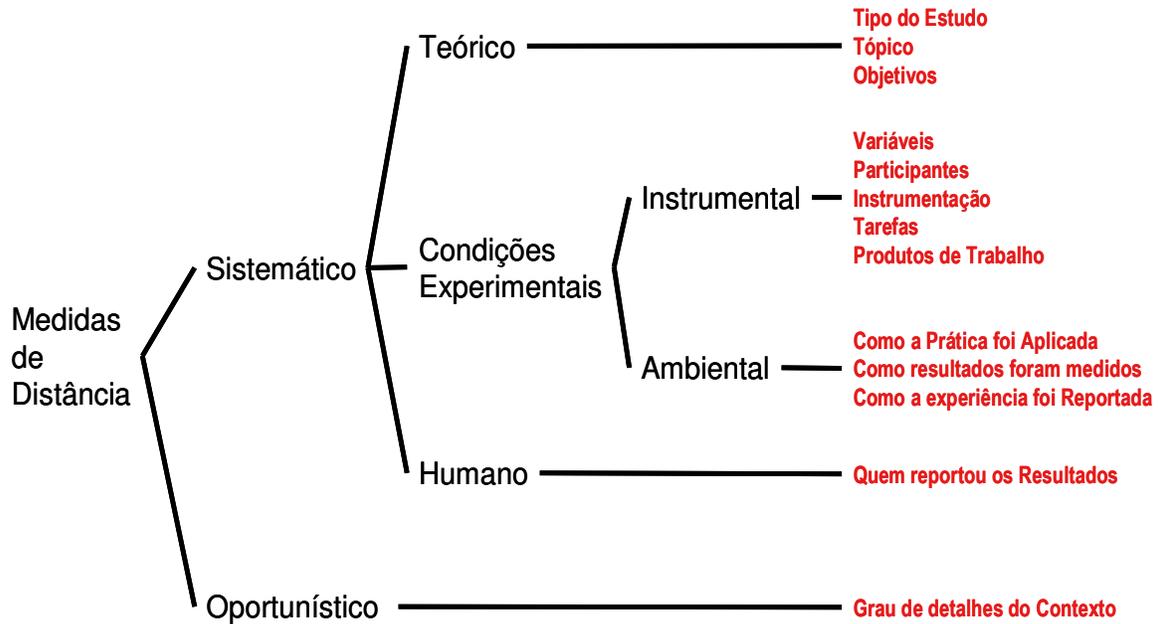
Definimos a árvore de distância mostrada na Figura 16. Os estudos podem ser aglomerados baseando-se nesta árvore da distância; os conceitos utilizados na figura são os seguintes.

- ✓ Sistemático: características do estudo que podem ter influência nos resultados e podem ser controlados pelos experimentadores. Eles podem ser relacionados a:
  - Teoria: consiste na teoria usada para projetar o estudo experimental. A SecESE define os seguintes atributos: tipo de estudo, tópico e objetivos. Outros podem ser definidos pelo analista.
  - Condições experimentais: diferenças que podem vir da instrumentação ou das condições experimentais.
  - Humanas: diferenças nos resultados podem vir de diferentes habilidades dos experimentadores na execução do estudo experimental, ou das pessoas que estão relatando os resultados.
- ✓ Oportunísticas: características do estudo que podem ter influência nos resultados, mas não foram controladas pelos experimentadores ou não foram relatadas nos artigos.

Existem ainda vários componentes e variáveis que não são descritos aqui; tentamos definir um conjunto de componentes que seria viável e mantivesse a simplicidade no processo. Nesta árvore, se dois contextos são iguais no Nível Zero eles serão iguais em todos os outros níveis. No Nível 1 existem duas possibilidades: similaridade sistemática e similaridade oportunística. No Nível 2 o analista pode analisar as similaridades e diferenças Teóricas, das Condições Experimentais e das questões Humanas.

Para realizar a tarefa de agrupar os contextos dos resultados definimos como diretriz os seguintes passos básicos:

- 1) Colete todas as informações dos contextos de todos os resultados associados a um padrão encontrado nos dados;
- 2) Defina por qual critério da árvore de distância os contextos devem ser agrupados;



**Figura 16 - Árvore de Distância baseada nas Informações de Contexto**

- 3) Aglomere os contextos; Existem duas grandes famílias de algoritmos de aglomerações. Os algoritmos hierárquicos e os algoritmos não hierárquicos. Os algoritmos não hierárquicos requerem que os seus usuários especifiquem a priori o número de grupos a serem gerados. Porém, o nosso problema é determinar um conjunto consistente de grupos e pré-determinar o número destes grupos pode fazer com que o algoritmo de aglomeração execute fusões ou quebras desnecessárias produzindo agrupamentos artificiais. Necessitamos de um processo de aglomeração em que o número apropriado de grupos possa ser interativamente verificado pelo usuário. Este processo pode ser alcançado pela utilização interativa de um algoritmo

---

hierárquico de aglomeração. Este tipo de algoritmo permite que os usuários controlem parâmetros para determinar o número apropriado de grupos [62].

### 3.6.2.5 Estabelecer conclusões

Analise cada padrão, resultados associados, contextos e qualidade dos resultados para estabelecer conclusões sobre os padrões que estão sendo avaliados sob a perspectiva do propósito inicial da análise. Este propósito foi definido no início da análise (Seção 3.3.1). Como diretriz estabelecemos os seguintes passos a serem seguidos:

- 1) Defina um nível mínimo de qualidade aceitável, descarte os resultados de qualidade insatisfatória, ou use-os e deixe claro na conclusão a qualidade dos resultados encontrados.
- 2) Reavalie os resultados.
- 3) Estabeleça as Conclusões. No caso do propósito da análise ser suporte, aceita-se ou rejeita-se uma hipótese inicial. No caso de sumarizar, deve-se analisar e concluir baseado no que os resultados mostram. No caso de comparar, deve-se analisar a evolução dos resultados baseados nos contextos e qualidade dos estudos comparando os resultados atuais com os resultados da base. A resolução final vai sempre depender dos resultados, dos contextos nos quais estes resultados foram encontrados e da qualidade destes resultados. Definimos algumas possibilidades de conclusão para dois tipos mais comuns de propósitos de análise. Algumas situações que podem existir assim como as possibilidades de conclusões são descritas na Tabela 12. Note que o quadro estabelece as conclusões para resultados de boa qualidade, mas claramente pode-se estabelecer conclusões para resultados de qualidade ruim, mas sempre deixando claro quais são as limitações destas conclusões.

**Tabela 12 - Possibilidades de Conclusões**

<b>Propósito da Análise quando Envolve Conteúdo da Informação</b>			
<p>Sumarizar: A análise deve ser feita comparando somente os resultados entre si. Resume-se o que foi encontrado baseando-se nos resultados, contexto e qualidade.</p> <p>Suporte: A análise deve ser feita comparando os Resultados com a Hipótese Inicial. Aceita-se ou rejeita-se a hipótese inicial a depender dos resultados, contextos e qualidade dos estudos.</p> <p>Comparação: A análise deve ser feita comparando os resultados atuais com os resultados da base. Analisa-se a evolução dos resultados baseados nos contextos e qualidade dos estudos.</p>			
<b>Resultados</b>	<b>Contextos</b>	<b>Qualidade dos Resultados</b>	<b>Conclusões</b>
Dois ou Mais resultados Concordantes	Iguais	Boa	Confirma e fortalece resultado
	Disjuntos	Boa	Generalização
	Interseccionais ou Contidos	Boa	Confirmação, Generalização ou limitação sendo que depende do grau de intersecção entre os contextos e do tipo de resultado.
Dois ou Mais resultados Discordantes	Iguais	Boa	Ou os contextos não foram suficientemente detalhados ou existem variáveis independentes desconhecidas ou abordadas em apenas alguns dos estudos.
	Disjuntos	Boa	Os resultados são dependentes do contexto. E as conclusões podem limitar os contextos nos quais os resultados podem ser reconhecidos como válidos.
	Interseccionais ou Contidos	Boa	Os resultados são dependentes do contexto, mas, existem pistas do que pode ter mudado o resultado.
Misto entre Concordantes e Discordantes	Iguais	Boa	Ou os contextos não foram suficientemente detalhados ou existem variáveis independentes desconhecidas ou abordadas em apenas alguns dos estudos.
	Disjuntos	Boa	Os resultados são dependentes do contexto. E as conclusões podem limitar os contextos nos quais os resultados podem ser reconhecidos como válidos.
	Interseccionais ou Contidos	Boa	Os resultados são dependentes do contexto, mas, existem pistas do que pode ter mudado o resultado.

---

4) Avalie as conclusões. A avaliação das conclusões estima a adequação da conclusão em relação aos critérios estabelecidos para um processo de descoberta de conhecimento. Por isso, a participação do especialista do domínio é fundamental, pois é ele quem detém as informações necessárias para verificar o interessantismo das conclusões obtidas (o quão interessante é a conclusão) [52]. Assim, o especialista no domínio pode mensurar o interessantismo do conhecimento descoberto, que é comumente caracterizado por critérios como [127]:

- **Validade:** indica o grau de certeza que se tem sobre o padrão encontrado, ou seja, a taxa de acerto obtida nos testes realizados, medida por critérios qualitativos ou estatísticos; esta validade também pode ser estabelecida pela avaliação de qualidade realizada durante este processo;
- **Utilidade:** indica a facilidade ou o potencial de utilização do conhecimento extraído em sistemas de aplicação ou de tomadas de decisão;
- **Novidade:** indica grau de interesse ou surpresa revelado pela conclusão, enfatizando a característica inovadora do mesmo, baseando-se no conhecimento prévio e nas expectativas do *especialista no domínio*;
- **Simplicidade:** indica a facilidade de compreensão do conhecimento extraído;
- **Generalidade:** indica o quão abrangente é o padrão obtido.

5) Formalize as conclusões: De acordo com o parecer do especialista no domínio, cada conclusão encontrada durante esta atividade pode ser consolidada como uma conclusão da análise e assimilada ou incorporada a sistemas de aplicação ou de tomada de decisões. Caso o especialista de domínio não fique satisfeito, a conclusão é descartada. Caso a conclusão seja consolidada, deve-se estabelecer e claramente relatar juntamente com as conclusões, quais foram os estudos usados como bases para esta conclusão, identificando os estudos favoráveis e não favoráveis à conclusão, os atenuantes que fizeram o analista não considerar as evidências não favoráveis, e quais as incertezas remanescentes.

### 3.7 Anotações

Muitos problemas aparecem durante a execução das atividades da SecESE e decisões de análise necessitam ser tomadas constantemente. O sucesso da análise dependerá do resultado das soluções e das decisões tomadas. A repetibilidade do processo também dependerá do registro destas decisões.

Portanto, os conhecimentos usado e novo constituem recursos que devem ser controlados eficientemente. A aquisição de conhecimento permite a melhor compreensão de seus processos e ajuda também na repetibilidade do processo da análise. Isto é difícil porque os analistas têm dificuldades em externar seu conhecimento. Assim, o processo da aquisição do conhecimento deve ser integrado ao processo da análise do formulário para minimizar o impacto na rotina normal do trabalho e do esforço de registrar o conhecimento [91]. O modelo utilizado aqui foi adaptado de Montoni [91].

As anotações têm como objetivo facilitar a coleta das informações importantes na condução de uma análise. Durante a execução de cada uma das atividades o analista deve anotar: decisões, idéias, lições aprendidas, dúvidas, melhoria do processo, conhecimento novo, comentários, conclusões, e tudo mais que seja necessário registrar. Definimos uma maneira estruturada de recolher esta informação, descrita na Tabela 13.

Além dos benefícios de repetibilidade expostos anteriormente, a documentação das anotações descritas na Tabela 13 traz outros benefícios, a saber:

- Documenta a informação, disponibilizando-a para a comunidade interessada.
- Novatos na área podem reusar o conhecimento de colaboradores experientes;
- Aumentam a produtividade, a eficiência e efetividade da análise dos resultados de um processo da extração do conhecimento, devido ao acesso ao conhecimento prévio para executar as atividades;
- Aumenta a confiabilidade do modelo, já que a informação sobre as decisões, suposições, idéias recolhidas durante o processo foi coletada e está disponível juntamente com a base de dados estruturada.

**Tabela 13 - Estrutura para cada tipo de anotação.**

<b>Tipo de Anotação</b>	<b>Estrutura da Anotação</b>
Decisões	Título
	Descrição
	Decisão
	Justificativa da Decisão:
Lições Aprendidas	Título
	Descrição
	Causa do Problema:
	Conseqüências do Fato
	Solução
	Resultados
Idéia	Título
	Descrição
	Fatos que suportam a idéia
	Situações nas qual a idéia foi aplicada e os resultados
Dúvida	Título
	Descrição
	Origem
Conhecimento novo sobre Domínio, Estudos Experimentais, Análise, Processo de Coleta, análise de dados etc.	Título
	Descrição

### 3.8 Mantendo uma Base de Resultados

Depois que uma análise é realizada, é possível que os seguintes tipos de resultados existam numa base estruturada de resultados.

- Resultados Primários:
  - Resultados obtidos em um estudo primário, descrito em um artigo. Um resultado primário é derivado de uma análise em um estudo experimental, ou de várias análises de alguns estudos experimentais em uma família. Este é um conhecimento que é geralmente restrito a um contexto específico. A confiança neste resultado está diretamente ligada à confiança dos resultados apresentados nos artigos e à qualidade medida na fase de avaliação da qualidade dos artigos.

- Resultados Secundários ou Conclusões:
  - Conclusões formuladas pela análise de resultados primários. Estes resultados são apoiados por mais de um estudo experimental. A confiança nestes resultados está diretamente ligada aos executores do processo, ao analista, ao especialista em experimentação, especialista no domínio e, principalmente, à qualidade dos resultados primários que suportam estes resultados.

Como o processo de extração A SecESE visa extrair todos os resultados de cada documento científico, então o artigo já coletado na base de dados pode ser reutilizado. Porém, para cada nova análise, deve-se sempre ler o artigo novamente, já que é importante a compreensão do artigo para que a análise seja feita conscientemente. A vantagem é que o artigo para ser reutilizado só precisa ser adicionado à análise e não precisa mais passar pelo processo de extração. Assim, à medida que algumas análises são realizadas, a base de conhecimento vai sendo expandida. Um conjunto maior de artigos é disponibilizado para novas análises. Também, já que existe uma maneira estruturada de coletar as informações, não é necessário que todas as extrações de informações dos artigos sejam feitas por uma só pessoa; isto facilita a expansão da base.

## Capítulo 4

### Apoio de Ferramentas

#### 4.1 Introdução

Para auxiliar a execução da SecESE, foi necessário o desenvolvimento e o uso de ferramentas para a simplificação, sistematização e padronização do processo de coleta das informações. Como mostrado na Figura 17, a SecESE requer diferentes tipos de apoio durante as suas quatro atividades. Uma ferramenta foi desenvolvida para a coleta das informações de análise e das informações coletadas dos artigos. Esta ferramenta serve de apoio às atividades de planejamento de análise, avaliação da qualidade e extração das informações, assim como gera os arquivos de entrada para outras duas ferramentas de mineração de dados que foram selecionadas para a análise de resultados: a *Treemap* e a HCE, ambas desenvolvidas na Universidade de Maryland<sup>3</sup> e de uso livre [18, 62]. Este capítulo descreve estas ferramentas e utiliza um estudo de viabilidade para demonstrar o uso destas durante a execução da SecESE.

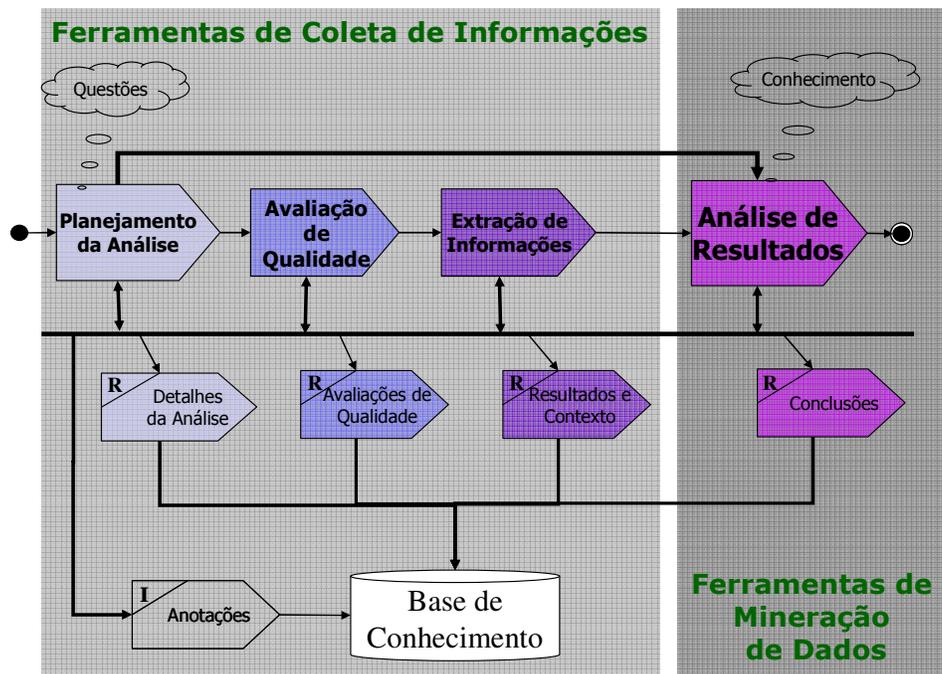


Figura 17 – Apoio de Ferramentas para a SecESE

<sup>3</sup> <http://www.cs.umd.edu/hcil/>

## **4.2 Coleta de informações - Ferramenta InfoESE**

Desenvolvemos uma ferramenta livre para dar apoio à coleta de informações de artigos. A ferramenta foi denominada InfoESE e seu principal objetivo é padronizar a coleta das informações definidas na SecESE. A ferramenta armazena as informações em uma base de dados e auxilia na formalização e padronização da coleta de informações, de forma a facilitar a análise das informações. Deve-se ressaltar que o foco desta tese não é o desenvolvimento do ferramental, mas o desenvolvimento e avaliação da SecESE. Por esta razão, a versão atual da InfoESE contém apenas o mínimo necessário para apoiar a coleta de informações definidas pela versão atual da SecESE para as atividades de planejamento da análise, avaliação da qualidade e extração das informações. Outras funcionalidades devem ser desenvolvidas para tornar a ferramenta mais efetiva principalmente na extração automática das informações dos artigos e na integração com as ferramentas de mineração de dados que serão usadas na atividade de análise de resultados. Nas seções que se seguem são mostradas a arquitetura e as funcionalidades da ferramenta.

### **4.2.1 Arquitetura e Modelo de Dados da Ferramenta InfoESE**

A ferramenta foi desenvolvida somente para a avaliação inicial da SecESE; foi desenvolvida no Access por questões de praticidade. O modelo de dados desenvolvido para esta ferramenta pode porém ser reaproveitado para desenvolvimento de uma ferramenta em Java com arquitetura apropriada para a incorporação das funcionalidades de visualização num ambiente integrado. Esta seção apresenta o diagrama de entidade e relacionamento na Figura 18 (o dicionário de dados é mostrado no Apêndice D); a próxima seção contém uma descrição mais detalhada das funcionalidades da ferramenta.

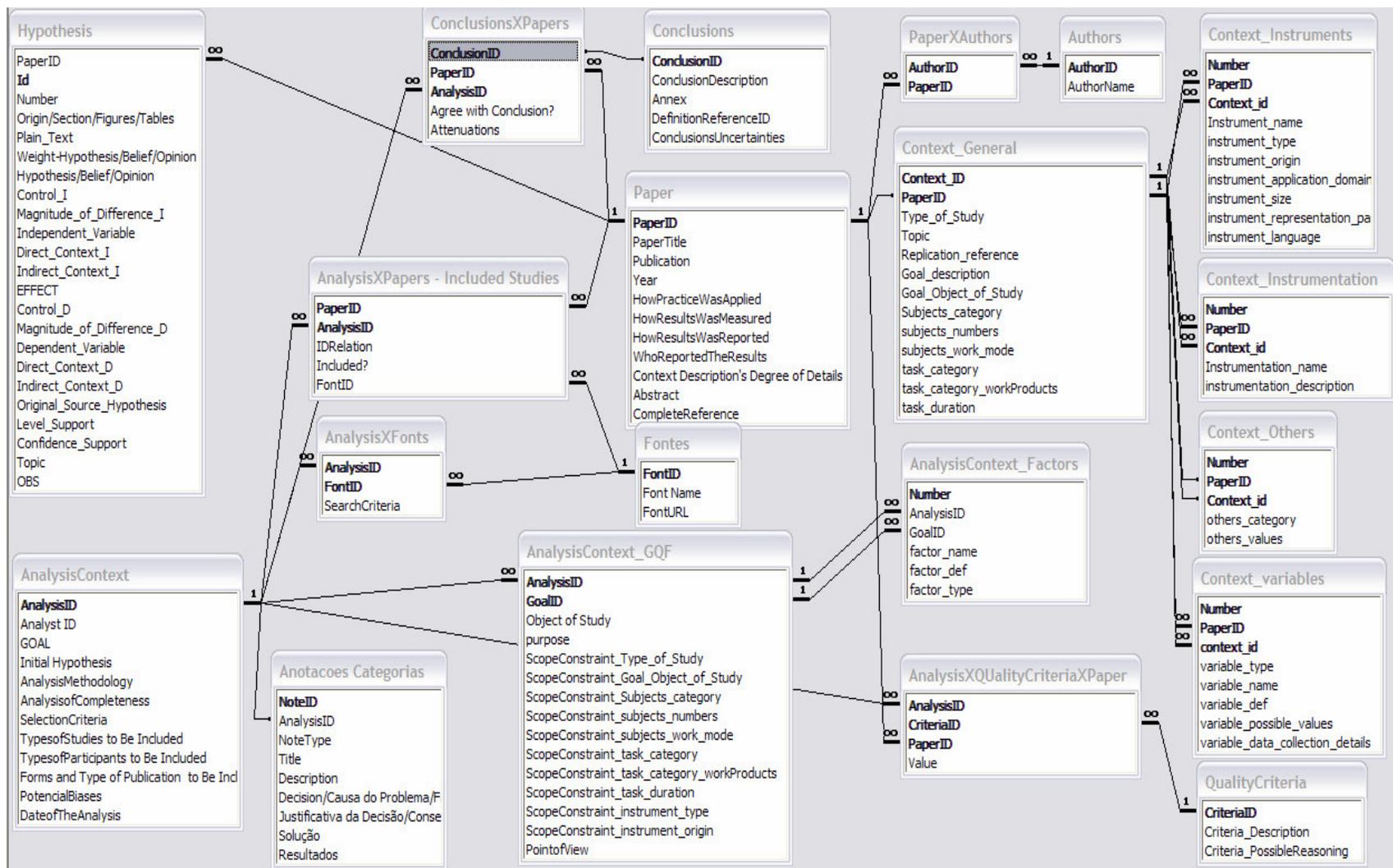


Figura 18 - Modelos de Dados da Ferramenta – InfoESE

## 4.2.2 Funcionalidades da Ferramenta InfoESE

O desenvolvimento da InfoESE foi realizado seguindo a abordagem SecESE (Figura 11); assim, as funcionalidades da ferramenta estão relacionadas às atividades de planejamento, avaliação da qualidade, extração de informações e análise dos resultados, como se segue.

### Planejamento

- Coletar informações do Plano de Análise, estas informações são armazenadas como detalhadas na Seção 3.3 e as Figura 19 e Figura 20 e mostram como as informações sobre o planejamento são coletadas para a base de dados.
- Inserir artigos na Análise
  - Uma vez que os artigos são selecionados na fase de análise e que a análise é criada no contexto da ferramenta, o usuário pode incluir os artigos na ferramenta e aloca-los à análise. Nesta fase somente as informações gerais dos artigos são coletadas (Figura 21). As seguintes informações são coletadas para cada artigo:
    - Título do Artigo
    - Publicação
    - Ano de Publicação
    - Resumo
    - Referência Completa do Artigo
    - Autores

### Extração de Informações

- Na extração das informações, informações mais detalhadas sobre o artigo podem ser coletadas, assim como os resultados coletados dos artigos e também as informações de contexto associadas a estes resultados. Três funcionalidades principais foram desenvolvidas para a extração das informações.
- Coletar Detalhes dos Artigos
  - Na coleta dos detalhes as informações sobre a qualidade dos artigos usando a mesma tela da Figura 21, completando então as informações que não estavam disponíveis no planejamento da análise como: como a prática foi aplicada, como os resultados foram medidos, como os resultados foram relatados, quem reportou os resultados e o grau de descrição das informações de contexto.
- Coletar Detalhes de Contexto.

---

○ A coleta das informações de contexto foi dividida em várias telas; as informações coletadas são detalhadas na Seção 3.5.4. A Figura 22 mostra a tela para coletar o que denominamos informações gerais sobre o contexto do estudo no artigo. A Figura 23 mostra a coleta das variáveis dependentes e independentes do estudo. A Figura 24 mostra a coleta dos detalhes da instrumentação. A coleta das informações sobre os produtos do trabalho é mostrada na Figura 25. Outras informações que o analista necessita coletar podem ser inseridas na base de dados como mostrado na Figura 26; neste estudo de viabilidade, era importante coletar a informação de se os defeitos analisados em cada programa foram semeados ou não.

- Coletar Resultados

- Na coleta de resultados as informações são coletadas como descritas na Seção 3.5.3, a Figura 27 mostra como esta coleta foi implementada.

## Análise

- Formalizar Resultados

- A primeira atividade da atividade de análise é formalizar os resultados coletados nos artigos de acordo com o processo descrito na Seção 3.6.1. A Figura 28 mostra como coletamos estas formalizações.

- Criar arquivos de entrada para as ferramentas de Visualização

- Para auxiliar o uso das ferramentas de apoio que serão descritas nas Seções 4.3 e 4.4 criamos algumas funcionalidades que auxiliam a criação dos arquivos de entrada para estas ferramentas.

- Coletar Conclusões da Análise.

- Na coleta das conclusões da análise (Figura 29), cada conclusão é associada aos artigos que suportam aquela conclusão; o analista deve declarar se aquele artigo suporta ou refuta aquela conclusão da análise.

Ao final do processo o analista terá uma base formalizada de resultados, contendo: detalhes da análise realizada, informações sobre os artigos e todos os resultados e informações de contexto associados a cada artigo que foi incluído na análise. Com esta base em mãos pode-se executar a análise dos resultados e das descrições do contexto.

Define Analysis Context		Analysis Goals	Papers	Register Contexts and Hypotheses	Run Analysis
<b>AnalysisID</b>	Testing Papers	<b>DateofTheAnalysis</b>	August 2006		
<b>Analyst ID</b>	Daniela Cruzes				
<b>Main GOAL</b>	<input type="checkbox"/> Analyze the information extraction technique to understand it with respect to completeness and consistency on extracting results and context from empirical studies papers on Testing Techniques from the viewpoint of Researchers in Software Engineering <input type="checkbox"/> Compare results with "Reviewing 25 years of Testing Technique Experiments"				
<b>Initial Hypothesis</b>	We will find the same conclusions they found.				
<b>AnalysisMethodology</b>	Take 12 of the papers and use our methodology to run the analysis.				
<b>AnalysisofCompleteness</b>					
<b>SelectionCriteria</b>	Papers related to functional Data Flow, control flow, random and mutation testing techniques included on the paper.				
<b>TypesofStudies to Be Included</b>	n/a				
<b>TypesofParticipants to Be Included</b>	n/a				
<b>Forms and Type of Publication to Be Included</b>	n/a				
<b>PotencialBiases</b>	The same she had on her paper... Is limited to her analysis.				

Figura 19 – Ferramenta InfoESE: Coleta de Informações de Análise

Define Analysis Context | Analysis Goals | Papers | Register Contexts and Hypotheses | Run Analysis

GoalID

Analyze **Object of Study** Reports  in order to **purpose** Explore

in respect to

Focus		focus_name	focus_def
▶		Error Distribution	0
*			0

and

Other Factors		factor_name	factor_def
▶		Error Distribution	0
		Errors Category	0
		Interface Errors Type	0

within the follow Scope Constraint

<b>ScopeConstraint_Type_of_Study</b> Case Study <input type="text"/>	<b>ScopeConstraint_Goal_Object_of_Study</b> <input type="text"/>
<b>ScopeConstraint_Subjects_category</b> Professionals <input type="text"/>	<b>ScopeConstraint_subjects_work_mode</b> <input type="text"/>
<b>ScopeConstraint_task_category</b> <input type="text"/>	<b>ScopeConstraint_task_category_workProducts</b> <input type="text"/>
<b>ScopeConstraint_instrument_type</b> <input type="text"/>	<b>ScopeConstraint_instrument_origin</b> <input type="text"/>

Figura 20 – Ferramenta InfoESE: Coletando Objetivos de Análise

Define Analysis Context | Analysis Goals | Papers | Register Contexts and Hypotheses | Run Analysis

PaperID: 23 | PaperTitle: The Cost of Data Flow Testing: An Empirical study

Navigation: [Previous] [Next] [Home] [End]

Paper Details | General Context | Variables | Instrumentation | Work Products | Other Information | Hypotheses | Hypotheses Formalization

**PaperTitle**  
The Cost of Data Flow Testing: An Empirical study

**Publication**  
IEEE Transactions on Software Engineering

Year	HowPracticeWasApplied	HowResultsWasMeasured
1990	4	5

HowResultsWasReported	WhoReportedTheResults	Context Description's Degree o
5	2	4

**Abstract**  
Abstract-A family of test data adequacy criteria employing data flow information has been previously proposed, and theoretical complexity analysis performed. This paper describes an empirical study to determine

**CompleteReference**  
Weyuker, E. J. 1990. The cost of data flow testing: An empirical study. IEEE Transactions on Software, Engineering 16(2): 121-128.

Figura 21 – Ferramenta InfoESE: Coleta de Informações Gerais sobre os Artigos

Define Analysis Context		Analysis Goals	Papers	Register Contexts and Hypotheses	Run Analysis		
PaperID	PaperTitle				Context_ID		
23	The Cost of Data Flow Testing: An Empirical study				1		
<input type="button" value="⏪"/> <input type="button" value="⏩"/> <input type="button" value="⏴"/> <input type="button" value="⏵"/>				Replication_reference			
Paper Details	General Context	Variables	Instrumentation	Work Products	Other Information	Hypotheses	Hypotheses Formalization
<b>Type of Study</b> Controlled Experiment							
<b>Description of the Topic</b> D.2.5 Testing and Debugging							
<b>Goal</b>							
Goal_description Analyze test sets for the purpose of analyse for its usefulness in detecting errors and evaluating the comprehensiveness. Investigate the costs of using various data flow testing techniques in practice. Confirm the intuition and informal experience that our family of software testing criteria based on the program's data flow characteristics is				Goal_Object_of_Study Software Engineering Techniques			
<b>Subjects</b>							
Category Graduate Students		Number of Subjects 4		Work Mode Individual			
<b>TASK</b>							
Category Create		Working Products Set of Test Case		Duration			

Figura 22 - Ferramenta InfoESE – Coleta de Informações gerais do Contexto dos Estudos

Define Analysis Context | Analysis Goals | Papers | Register Contexts and Hypotheses | Run Analysis

PaperID: 23 | PaperTitle: The Cost of Data Flow Testing: An Empirical study

Navigation: [Home] [Previous] [Next] [End]

Paper Details | General Context | Variables | Instrumentation | Work Products | Other Information | Hypotheses | Hypotheses Formalization

Number	Type	Name	Possible Values	Data Collection Details
1	Independent	Size		number of decisions statements
2	Independent	Test Sets		
3	Independent	Testing Techniques	all-c-uses, all-p-uses, all-uses, and all-dupaths	
▶ 4	Dependent	▼ Criteria Compliance		The least squares line: $t = ad + P$ , where t is the number of test cases sufficient to satisfy the given criterion for the subject program, and d is the number of decision statements in the subject program
5	Dependent	Number of Test Cases		

Figura 23 - Ferramenta InfoESE – Coleta de Variáveis Dependentes e Independentes.

Define Analysis Context | Analysis Goals | Papers | Register Contexts and Hypotheses | Run Analysis

PaperID: 23 | PaperTitle: The Cost of Data Flow Testing: An Empirical study

Navigation: [Previous] [Next] [Home] [End]

Paper Details | General Context | Variables | Instrumentation | Work Products | Other Information | Hypotheses | Hypotheses Formalization

	Number	Instrumentation_name	instrumentation_description
▶	1	Asset	Automate the application of the techniques. Input 29 Subprograms from Software Tools in Pascal (Kernighan and Plauger Suite) with five or more decision statements. It generates practically the whole output information
*	0		

Figura 24 - Ferramenta InfoESE – Coleta de Informações sobre Instrumentação

Define Analysis Context | Analysis Goals | Papers | Register Contexts and Hypotheses | Run Analysis

PaperID: 23 | PaperTitle: The Cost of Data Flow Testing: An Empirical study

Navigation: [Home] [Previous] [Next] [End]

Paper Details | General Context | Variables | Instrumentation | Work Products | Other Information | Hypotheses | Hypotheses Formalization

	Number	Name	Type	Origin	Application Domain	Size	Representation Paradigm	Language
▶	1	Kernighan	Code	Commercial	Software Library Tool		Structured	Pascal
*	0							

**Figura 25 - Ferramenta InfoESE – Coleta de Informações sobre os Produtos de Trabalho**

Define Analysis Context | Analysis Goals | Papers | Register Contexts and Hypotheses | Run Analysis

PaperID	PaperTitle
23	The Cost of Data Flow Testing: An Empirical study

⏪ ⏩ ⏴ ⏵

Paper Details | General Context | Variables | Instrumentation | Work Products | Other Information | Hypotheses | Hypotheses Formalization

	Number	Category	Value
▶	1	Faults Seeded?	No
*	0		

Figura 26 - Ferramenta InfoESE – Coleta de Outras Informações

Define Analysis Context | Analysis Goals | Papers | Register Contexts and Hypotheses | Run Analysis

PaperID: 25 | PaperTitle: An experimental comparison of the effectiveness of branch testing and data flow testing.

Navigation: [Previous] [Next] [Home] [End]

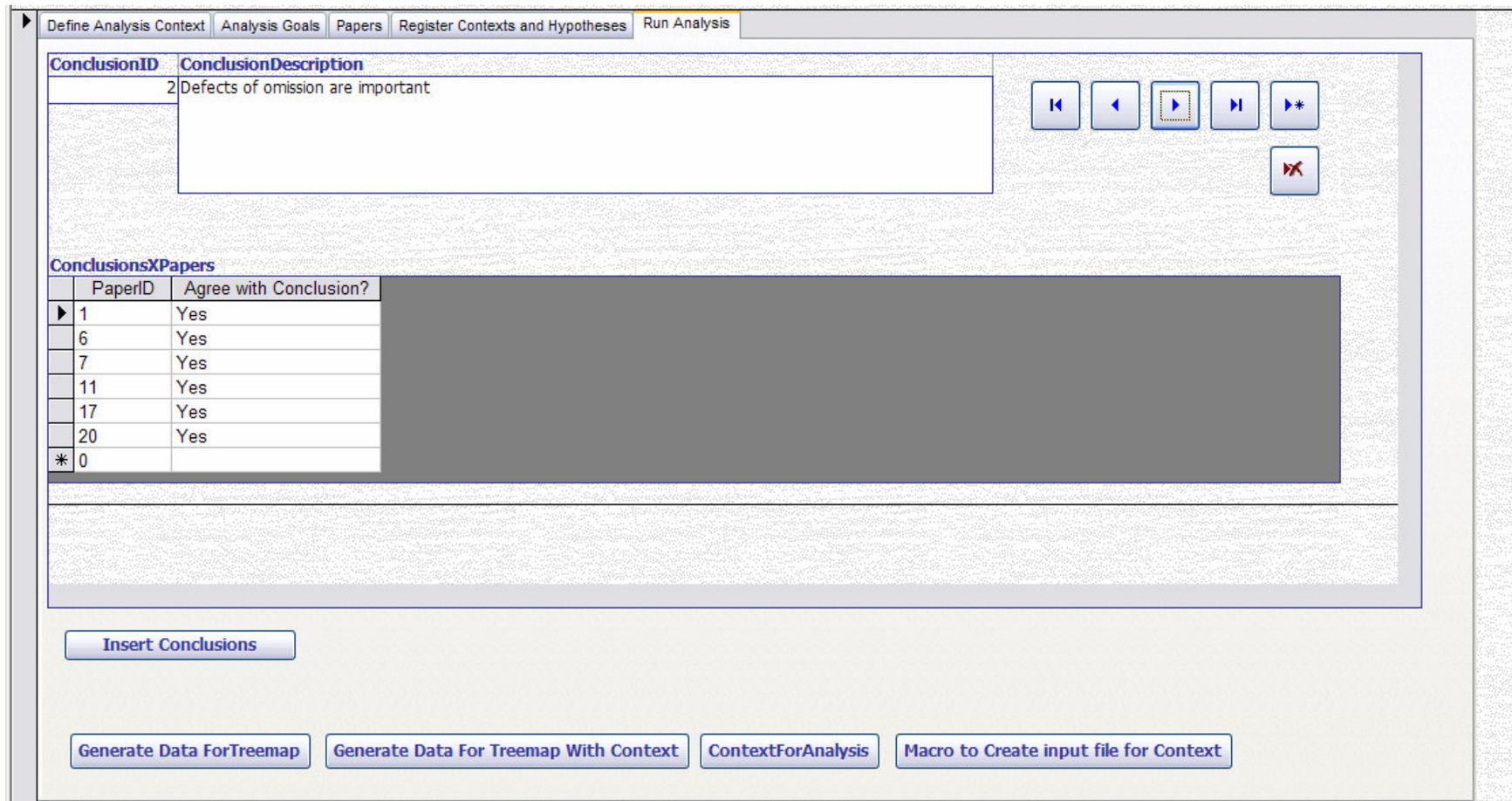
Paper Details | General Context | Variables | Instrumentation | Work Products | Other Information | Hypotheses | Hypotheses Formalization

	Id	Number	Plain_Text	[Origin/Section/f	Hypothesis/Be	Level_Support	OBS
▶	753	1	Five of the nine subjects, all-uses was more effective than all-edges at 99% confidence;	Experimental Results	Hypothesis	significantly positive	
	754	2	for six subjects, all-uses was more effective than the null criterion;	Experimental Results	Hypothesis	significantly positive	
	755	3	For five subjects all-edges was more effective than null.	Experimental Results	Hypothesis	significantly positive	

Figura 27 - Ferramenta InfoESE – Coleta de Resultados

PaperID	PaperTitle		Context_ID																										
25	An experimental comparison of the effectiveness of branch testing and data flow testing.		1																										
<div style="display: flex; justify-content: space-between;"> <span>Replication_reference</span> </div>																													
<div style="display: flex; justify-content: space-between;"> <span>Paper Details</span> <span>General Context</span> <span>Variables</span> <span>Instrumentation</span> <span>Work Products</span> <span>Other Information</span> <span>Hypotheses</span> <span>Hypotheses Formalization</span> </div>																													
<table border="1"> <thead> <tr> <th>Hypothesis Id</th> <th>Number</th> <th>Hypothesis/Belief/Opinion</th> </tr> </thead> <tbody> <tr> <td>753</td> <td>1</td> <td>Hypothesis</td> </tr> </tbody> </table> <p><b>Plain_Text</b> Five of the nine subjects, all-uses was more effective than all-edges at 99% confidence;</p> <div style="border: 1px solid gray; padding: 5px;"> <p><b>Independent Side</b></p> <table border="1"> <tr><td>Control_I</td><td></td></tr> <tr><td>Magnitude_of_Difference_I</td><td></td></tr> <tr><td>Independent_Variable</td><td>Effectiveness</td></tr> <tr><td>Direct_Context_I</td><td>all-uses</td></tr> <tr><td>Indirect_Context_I</td><td>in five of the nine subjects</td></tr> </table> <p><b>EFFECT</b> Greater</p> <p><b>Dependent Side</b></p> <table border="1"> <tr><td>Control_D</td><td></td></tr> <tr><td>Magnitude_of_Difference_D</td><td></td></tr> <tr><td>Dependent_Variable</td><td>Effectiveness</td></tr> <tr><td>Direct_Context_D</td><td>all-edges</td></tr> <tr><td>Indirect_Context_D</td><td></td></tr> </table> </div>				Hypothesis Id	Number	Hypothesis/Belief/Opinion	753	1	Hypothesis	Control_I		Magnitude_of_Difference_I		Independent_Variable	Effectiveness	Direct_Context_I	all-uses	Indirect_Context_I	in five of the nine subjects	Control_D		Magnitude_of_Difference_D		Dependent_Variable	Effectiveness	Direct_Context_D	all-edges	Indirect_Context_D	
Hypothesis Id	Number	Hypothesis/Belief/Opinion																											
753	1	Hypothesis																											
Control_I																													
Magnitude_of_Difference_I																													
Independent_Variable	Effectiveness																												
Direct_Context_I	all-uses																												
Indirect_Context_I	in five of the nine subjects																												
Control_D																													
Magnitude_of_Difference_D																													
Dependent_Variable	Effectiveness																												
Direct_Context_D	all-edges																												
Indirect_Context_D																													
Record: <span>⏪</span> <span>⏩</span> 1 of 23 <span>⏴</span> <span>⏵</span> <span>⏶</span> <span>⏷</span>																													

Figura 28 - Ferramenta InfoESE – Formalização de Resultados



**Figura 29 – Ferramenta InfoESE – Coleta de Conclusões da Análise**

---

### **4.3 Análise dos Resultados com Exploração Visual de Dados**

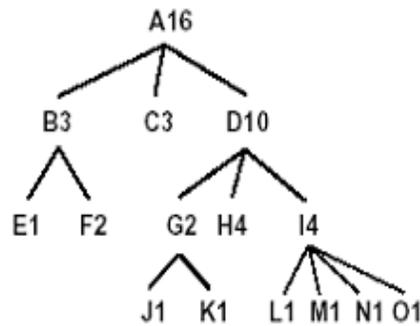
A visualização de informações visa auxiliar o processo de análise e compreensão de um conjunto de dados usando representações gráficas manipuláveis destes dados. As técnicas de visualização de informações procuram representar graficamente dados de um determinado domínio de aplicação de modo que a representação visual gerada explore a capacidade de percepção humana para, a partir das relações espaciais exibidas, interpretar e compreender as informações apresentadas [33, 90].

No contexto da SecESE, uma das opções é o uso de uma técnica de exploração visual hierárquica de informações para explorar a base de resultados previamente construída. Estas técnicas permitem que se organize hierarquicamente, segundo um determinado critério, os dados sendo explorados. Isto equivale a organizar as informações de acordo com atributos dinamicamente escolhidos pelo cientista que está explorando a base de resultados. A característica dinâmica deste tipo de ferramenta permite que um perito possa rapidamente chavear contextos de exploração pela modificação dos critérios hierárquicos usados para organizar os resultados sendo explorados.

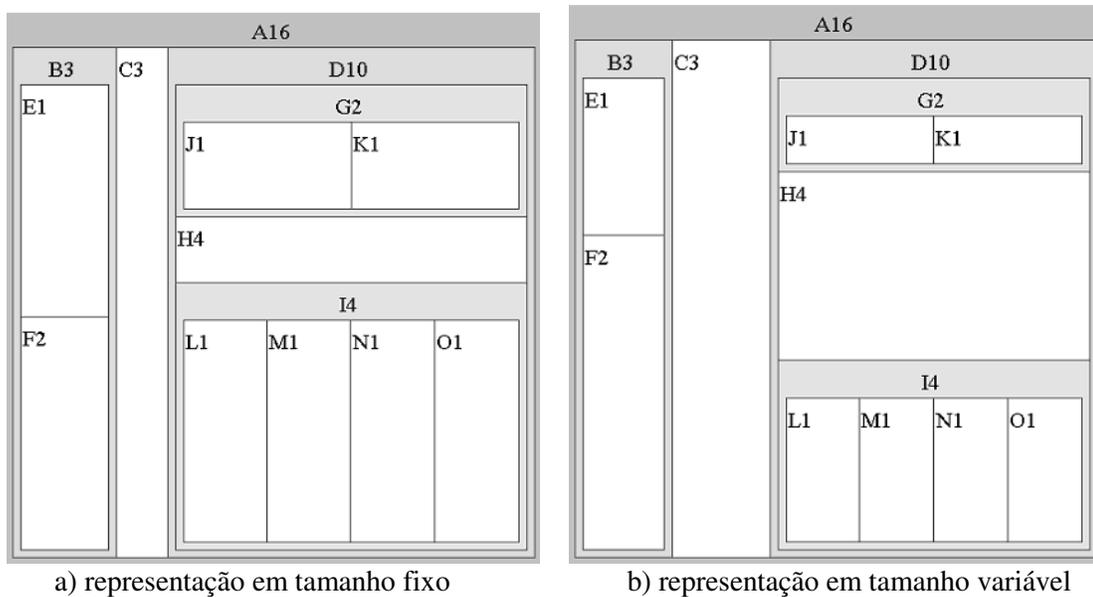
#### **4.3.1 A Técnica de Visualização para Exploração dos Resultados**

Selecionamos os Mapas em Árvores ou *Treemaps* para a exploração dos resultados por dois motivos principais: i) esta é uma técnica bastante difundida e intuitiva de visualização hierárquica de informações; ii) existe a disponibilidade de boas ferramentas de livre acesso que implementam esta técnica de visualização.

O Mapa em Árvore [18] é um método de visualização de estruturas hierárquicas que utiliza todo o espaço disponível na tela de visualização. Este método utiliza 100% do espaço disponível pelo mapeamento da hierarquia em regiões retangulares aninhadas. A Figura 31a apresenta a representação dos dados hierárquicos da árvore da Figura 30 como um mapa em árvore.



**Figura 30 – Desenho tradicional de uma hierarquia em uma árvore**



**Figura 31 – Representação do Mapa em Árvore**

Por utilizar todo o espaço disponível na sua representação, os mapas em árvore permitem a representação eficiente de grandes hierarquias. Além disto, os mapas em árvores podem mostrar atributos dos nós pelo tamanho e cor dos seus retângulos. O atributo visual "tamanho" é especialmente útil na representação de variáveis que podem ser decompostas hierarquicamente. Como mostrado na Figura 31b, os mapas em árvore podem ser usados para fazer com que os nós que contenham informações de maior importância sejam colocados em regiões maiores que aqueles de menor importância. Isto ajuda a mostrar padrões hierárquicos permitindo aos usuários comparar os tamanhos dos nós e das subárvores.

---

### 4.3.2 A Ferramenta para Exploração dos Resultados

A ferramenta escolhida para exploração hierárquica da base de resultados foi a “*Treemap*”. A *Treemap* é uma ferramenta para explorar mapas em árvore desenvolvida pelo Human-Computer Interaction Laboratory (HCIL) da Universidade de Maryland (<http://www.cs.umd.edu/hcil/>). A ferramenta foi escolhida por ter as seguintes características:

- É de acesso livre;
- Tem uma implementação estável e confiável;
- Tem um grande número de funcionalidades para exploração visual de dados, incluindo:
  1. Filtragem dinâmica de informações por meio de *widjets*;
  2. Fácil navegação visual;
  3. Mapeamento dinâmico entre atributos reais (dos dados) e atributos visuais (do mapa em árvore);
  4. Obtenção de detalhes sobre demanda;
- É intuitiva e de uso fácil;
- Importa e exporta dados de diversas fontes, o que facilitou a criação de uma exportação de dados compatível a partir da InfoESE.

A Figura 32 mostra um retrato da tela da ferramenta *Treemap*. A ferramenta é dividida em três grandes áreas. À esquerda, a ferramenta mostra um mapa em árvore navegável hierarquicamente e interativo. No canto superior direito, a ferramenta mostra o detalhe sob demanda do último registro de dados clicado no mapa em árvore. No canto inferior direito, a ferramenta apresenta seus controles de interação. Todos os controles da *Treemap* estão distribuídos em quatro abas: Principal, Legenda, Filtros, e Hierarquia. Na aba principal, os usuários podem selecionar um dos três algoritmos de desenho do mapa em árvore, quadrado, fatia e corta, e em tiras, dependendo de suas necessidades, assim como opções do tamanho das fontes e das opções de borda. A aba da legenda permite que os usuários atribuam os mapeamentos entre os atributos dos dados aos atributos visuais do mapa (etiqueta, tamanho e cor dos retângulos). A *Treemap* permite ainda agrupar os valores numéricos de um atributo e atribuir cores a cada grupo.

O usuário pode fazer três tipos de interações diretamente sobre a área do mapa. Ao clicar em um de seus retângulos, o usuário faz com que a ferramenta mostre os detalhes do registro associado com o retângulo e apresente o trajeto que vai desse retângulo até a raiz da hierarquia. Enquanto o analista move o mouse sobre o mapa, um pop-up amarelo mostra os valores dos atributos atribuídos prescritos para etiqueta, tamanho e cor do retângulo. Um clique duplo na borda de um grupo de retângulos faz o mapa descer para o seu nível na hierarquia, enquanto que um clique no botão direito do mouse realiza a operação inversa.

Na aba do filtro da área de controle, os usuários podem filtrar dados usando *widgets* que realizam consultas interativas e dinâmicas sobre os dados sendo analisados. *Widgets* são associados a todos os atributos categóricos e numéricos dos dados sendo analisados. Quando o atributo é categórico, os dados são filtrados através de barras de itens (que permitem a seleção de todos ou um valor em particular), botões de rádio (que permitem a seleção de um valor em particular) ou caixas de verificação (que permitem a seleção de múltiplos valores de interesse). Quando o atributo é numérico, os dados são filtrados por barras de intervalo que permitem a seleção de um intervalo de valores de interesse. Os itens cujos atributos estão fora da escala selecionada são apresentados em cinza ou removidos do mapa, conforme opção do usuário. O processo é dinâmico e com um nível de interação praticamente instantâneo entre seleção e atualização da tela visual.

#### **4.3.3 Exemplo de uso Ferramenta Treemap para Exploração da Base de Resultados.**

A utilização da ferramenta *Treemap* para explorar bases de resultados envolve os seguintes passos:

1. Os resultados de interesse devem ser importados para a ferramenta. Vários formatos são aceitos e a InfoESE exporta os resultados em um formato compatível.
2. O usuário deve indicar como os dados serão explorados visualmente, escolhendo os atributos que definirão a hierarquia e os atributos que serão mapeados visualmente.
3. Uma vez montada uma estrutura visual, o usuário pode explorar os resultados dentro daquele cenário por meio dos filtros e mecanismos de interação visual fornecidos pela

---

ferramenta; para cada nova conclusão (*insight*) o usuário deverá estabelecer suas conclusões como discutido na Seção 3.6.2.5 e salvar a estrutura visual que a produziu.

Considere a Figura 33 como exemplo. Esta figura mostra um conjunto de aproximadamente 200 resultados extraídos de 11 artigos. Eles estão agrupados hierarquicamente segundo o atributo “Como a prática do estudo foi aplicada” (Os valores possíveis estão descritos na Tabela 14), e coloridos segundo o atributo “PaperID”. Nenhum atributo está associado com o tamanho do retângulo.

Na Figura 32 todos os resultados estão igualmente organizados na tela, sem hierarquia, e fica então difícil de se decidir por onde começar a busca por padrões ou respostas para a análise. Já na Figura 33 o usuário pode, por exemplo, começar e direcionar a busca nos resultados de projetos pilotos ou projetos em produção. Sendo assim, a análise se reduz a 34 resultados em 4 artigos, facilitando assim uma análise em conjuntos de resultados grandes.

**Tabela 14- Critérios de Qualidade - Fator Como a prática foi aplicada.**

Critério	Descrição	Valor
Como a Prática foi Aplicada	Não é claro	0
	Um Experimento	3
	Uma série de experimentos	4
	Um projeto piloto	4
	Uma série de projetos pilotos	5
	Um projeto em produção	5
	Uma série de projetos em produção	7

Note que a ferramenta *Treemap* pode auxiliar a exploração dos resultados dos artigos associados à suas informações de contexto. Isto auxilia uma análise prévia dos contextos antes da atividade de análise de contexto (Figura 15). Porém, é importante que os contextos sejam analisados separadamente e em detalhe, já que o entendimento da abrangência e escopo dos resultados que estão sendo analisados e combinados depende da compreensão do contexto no qual estes resultados foram coletados. Neste cenário, é desejável que a utilização de uma ferramenta como a *Treemap* seja combinada com alguma ferramenta que ajude a agrupar os estudos por similaridade. A próxima seção trata esta questão.



Figura 32 - Ferramenta Treemap



## 4.4 Apoio de Ferramentas para Análise de Contextos

Dentro do processo de analisar resultados tem-se um especial destaque à análise de contextos, pois a explicação para similaridades e diferenças de resultados nos estudos quase sempre está explícita nas diferenças e similaridades dos contextos nos quais estes resultados foram extraídos. Há por exemplo, alguns casos em que os resultados de diferentes fontes são contraditórios e o analista deve estar ciente dos contextos em que cada um dele foi gerado para fazer a interpretação correta sobre as razões pelas quais eles divergem. Este processo pode ser realizado pela utilização interativa de um algoritmo hierárquico de aglomeração, ou HAC (*Hierarchical Agglomerative Clustering*). Este tipo de algoritmo permite que os usuários controlem parâmetros para determinar o número apropriado de grupos [62].

### 4.4.1 Aglomeração Hierárquica de Dados

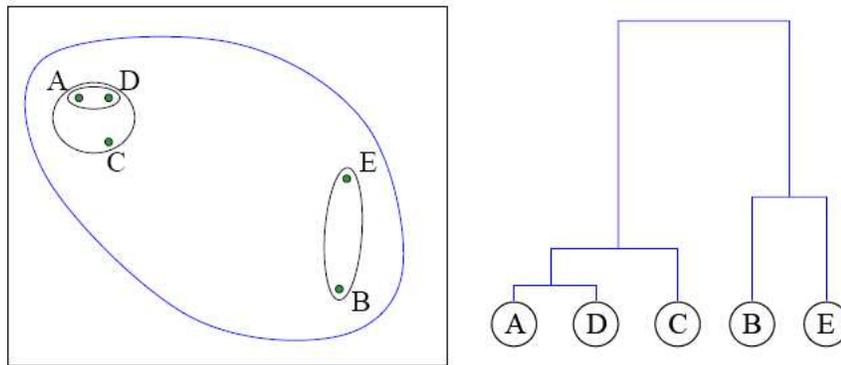
O algoritmo HAC é resumido como segue. Supomos que queremos aglomerar  $n$  itens de dados, e temos  $n * (n-1)/2$  valores de similaridade (ou distância) entre cada par possível de  $n$  itens de dados [62]:

1. Inicialmente, cada item de dado ocupa um grupo sozinho, portanto, no início são  $n$  grupos.
2. Encontre um par de grupos cujo valor de similaridade é maior e faça do par um novo grupo.
3. Atualize os valores de similaridade entre o novo grupo e os grupos remanescentes.
4. Passo 2 e 3 são aplicados  $n-1$  vezes até que reste somente um grupo de tamanho  $n$ .

Há muitas escolhas possíveis para atualizar os valores de similaridade da etapa 3. Entre eles, os mais comuns são: enlace completo, enlace médio, e enlace único. O enlace completo ajusta os valores da similaridade entre o grupo novo e os grupos restantes para ser o mínimo das similaridades entre cada membro do grupo novo e o resto. O enlace médio usa o valor médio da similaridade como o novo valor de similaridade. O enlace único usa o valor máximo.

Os resultados da aglomeração hierárquica são representados geralmente como dendrogramas. Um dendrograma é uma árvore binária, onde cada item de dados corresponde

a um nó terminal da árvore binária. A distância da raiz para uma sub-árvore indica a similaridade da sub-árvore - os nós altamente similares ou as sub-árvores têm os pontos de ligação que são mais distantes da raiz. Por exemplo, na Figura 34, a distância Euclidiana entre A e D são as menores entre todos os pares possíveis, e são fundidos como uma sub-árvore; a altura da sub-árvore é muito curta porque é muito similar nos termos da medida da similaridade/distância. Por outro lado, B e E não estão assim tão perto, a altura da sub-árvore correspondente é muito mais alta porque não são tão similares. Ainda na Figura 34, cinco pontos de dados (A, B, C, D, E) em um plano 2D são aglomerados, e o dendrograma (uma árvore binária) no lado direito mostra o resultado da aglomeração usando o enlace único e a distância Euclidiana. A altura de cada sub-árvore representa a distância entre os dois filhos.



**Figura 34 – Aglomeração Hierárquica e o dendrograma**

A representação por dendrogramas ajuda no entendimento de como podemos usá-los para interpretar os resultados obtidos em diferentes contextos. O que se necessita então é de um ferramental capaz de permitir o uso interativo de um HAC nas mesmas linhas definidas para a ferramenta de exploração hierárquica de informação apresentada anteriormente.

#### **4.4.2 Aglomeração com a ferramenta HCE**

A HCE (*Hierarchical Clustering Explorer*) [62] foi desenvolvida originalmente para a visualização interativa de resultados do agrupamento hierárquico de conjuntos de dados multidimensionais. A HCE foi desenvolvida no HCIL da Universidade de Maryland, o

mesmo laboratório que desenvolveu o *Treemap*. A ferramenta foi escolhida por ter a seguintes características:

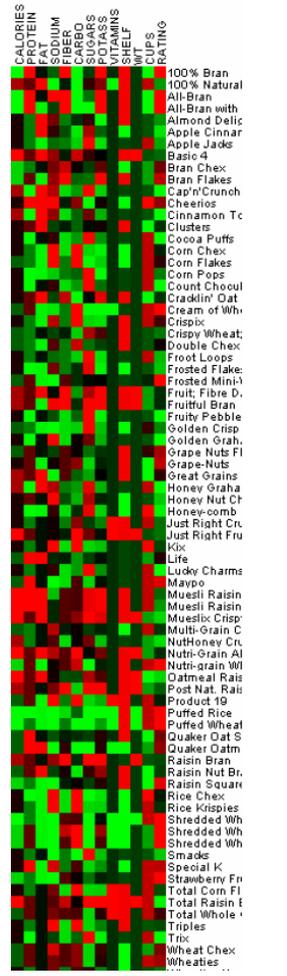
- É de acesso livre;
- Tem uma implementação estável e confiável;
- Importa e exporta dados de diversas fontes, o que facilitou a criação de uma exportação de dados compatível a partir da InfoESE.
- Além de implementar um HAC que constrói por aglomeração um modelo hierárquico descritivo dos dados que lhes são fornecidos (Figura 35);
- Possui os recursos de uma ferramenta de visualização de informação, é completamente interativa e usa uma cena visual para representar os agrupamentos que constrói.
- É útil a quem quer “visualizar” sua série de dados, “explorar” padrões interessantes de aglomeração, e “escolher” o modelo de aglomeração mais adequado à sua situação.

Esta seção descreve como o HCE utiliza visualização para auxiliar a compreensão de conjuntos de dados multidimensionais utilizando a exploração interativa de resultados de aglomerações hierárquicos, consultas dinâmicas e coordenação entre visões múltiplas.

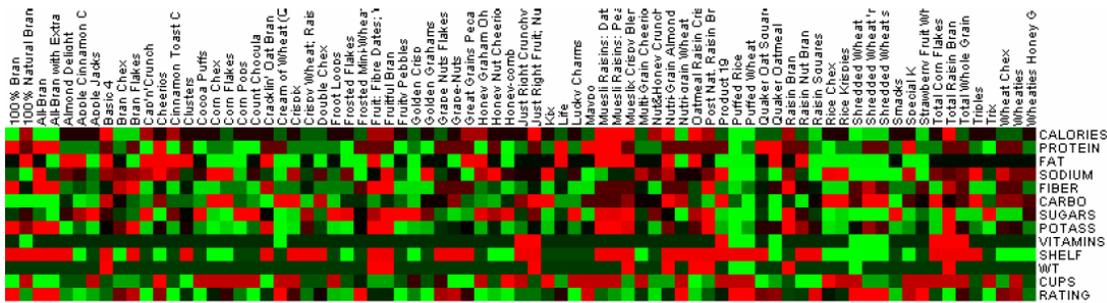
A HCE combina a apresentação de dendrogramas com um mosaico colorido descritivo dos dados sendo analisados. As séries de dados multidimensionais são representadas geralmente em uma tabela onde uma linha representa um item e uma coluna representa uma variável (ou uma dimensão). Por exemplo, Figura 35(a) mostra uma série de dados multidimensional pequena (77 linhas e 13 colunas) sobre a informação nutricional de cereais. Cada linha é um cereal, e cada coluna é um componente de nutrição. Uma representação gráfica deste conjunto de dados codifica cada valor na tabela de acordo com um esquema de mapeamento de cores. Esta representação gráfica de uma tabela é o que chamamos de “mosaico colorido”. Uma maneira típica de mostrar um mosaico colorido é manter a mesma disposição da tabela original e somente codificar cada célula (Figura 35(b)). Mesmo que esta disposição vertical seja uma representação natural, HCE usa uma disposição transposta (Figura 35(c)). Quando os analistas querem identificar pontos de interesse e compreender a distribuição dos dados, podem examinar as cores no mosaico.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	name	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	wt	cups	rating	
2	100% Bran	70	4	1	130	10	5	6	280	25	3	1	0.33	68.40287	
3	100% Natu	120	3	5	15	2	8	8	155	0	3	1	1	33.98388	
4	All Bran	70	4	1	280	9	7	5	320	25	3	1	0.33	59.42551	
5	All Bran w/	50	4	0	140	14	8	0	330	25	3	1	0.5	93.70491	
6	Almond Oa	110	2	2	200	1	14	8	70	25	3	1	0.75	34.36454	
7	Apple Crn	110	2	2	180	15	10.5	10	70	25	3	1	1	25.00954	
8	Apple Jack	110	2	0	125	1	11	14	30	25	2	1	1	33.17409	
9	Basic 4	130	3	2	210	2	18	8	100	25	3	1.33	0.75	37.03936	
10	Bran Chex	90	2	1	200	4	15	6	125	25	3	1	0.67	49.12025	
11	Bran Flak	90	3	0	210	5	13	5	190	25	3	1	0.67	53.31361	
12	Cap'n Crunch	120	1	2	220	0	12	12	35	25	2	1	0.75	18.04205	
13	Cheerios	110	6	2	290	2	17	1	105	25	1	1	1.25	50.725	
14	Cinnamon	120	1	3	210	0	13	9	45	25	2	1	0.75	19.82357	
15	Clusters	110	3	2	140	2	13	7	105	25	3	1	0.5	40.40021	
16	Cocoa Puf	110	1	1	180	0	12	13	55	25	2	1	1	22.73645	
17	Com Chex	110	2	0	280	0	22	3	25	25	1	1	1	41.44502	
18	Com Flak	100	2	0	290	1	21	2	35	25	1	1	1	45.56332	
19	Com Pop	110	1	0	90	1	13	12	20	25	2	1	1	35.78279	
20	Count Cho	110	1	1	180	0	12	13	65	25	2	1	1	22.39651	
21	Cracklin' C	110	3	3	140	4	10	7	160	25	3	1	0.5	40.44377	
22	Cream of V	100	3	0	80	1	21	0	-1	0	2	1	1	64.53362	
23	Crisps	110	2	0	220	1	21	3	30	25	3	1	1	46.89564	
24	Crispy Wh	100	2	1	140	2	11	10	120	25	3	1	0.75	31.1752	
25	Crunch Ch	100	2	0	190	1	19	5	60	25	3	1	0.75	44.33036	
26	Frost Loop	110	2	1	125	1	11	13	30	25	2	1	1	32.40759	
27	Frosted Fl	110	1	0	200	1	14	11	25	25	1	1	0.75	31.43587	
28	Frosted Wh	100	3	0	0	3	14	7	100	25	2	1	0.8	58.34514	
29	Fruit, Fibr	120	3	2	160	5	12	10	200	25	3	1.25	0.67	40.91705	
30	Fruiti Fib	120	3	0	240	5	14	12	190	25	3	1.33	0.67	41.01549	
31	Fruity Peb	110	1	1	135	0	13	12	25	25	2	1	0.75	28.02577	
32	Golden Cr	100	2	0	45	0	11	15	40	25	1	1	0.88	35.25244	
33	Golden Cr	110	1	1	280	0	15	9	45	25	2	1	0.75	23.80404	
34	Grape Nut	100	3	1	140	3	15	5	85	25	3	1	0.88	52.0769	
35	Grape Nut	110	3	0	170	3	17	3	90	25	3	1	0.25	53.37101	
36	Great Gra	120	3	3	75	3	13	4	100	25	3	1	0.33	45.31172	
37	Honey Gra	120	1	1	220	1	12	11	45	25	1	1	1	21.87128	
38	Honey Nut	110	3	1	260	1.5	11.5	10	90	25	1	1	0.75	31.07222	
39	Honey com	110	1	0	180	0	14	11	36	25	1	1	1.33	28.74241	
40	Just Right	110	2	1	170	1	17	6	60	100	3	1	1	1	36.52339
41	Just Right	140	3	1	170	2	20	9	95	100	3	1.3	0.75	36.47151	
42	Kix	110	2	1	260	0	21	3	40	25	2	1	1.5	39.24111	
43	Life	100	4	2	150	2	12	6	95	25	2	1	0.67	45.23007	
44	Lucky Cha	110	2	1	180	0	12	12	55	25	2	1	1	26.73452	
45	Maypo	100	4	1	0	0	16	3	55	25	2	1	1	54.85032	
46	Muesli Rai	150	4	3	95	3	16	11	170	25	3	1	1	37.13696	
47	Muesli Rai	150	4	3	150	3	16	11	170	25	3	1	1	54.13377	
48	Mueslix Cr	160	3	2	150	3	17	13	160	25	3	1.5	0.67	30.31265	
49	Multi-Gran	100	2	1	220	2	15	6	90	25	1	1	1	40.10587	
50	Nut&Hone	120	2	1	190	0	15	9	40	25	2	1	0.67	29.92429	
51	Nutri-Gran	140	3	2	220	3	21	7	130	25	3	1.33	0.67	40.89232	
52	Nutri-gran	90	3	0	170	3	18	2	90	25	3	1	1	59.64284	
53	Oatmeal R	130	3	2	170	1.5	13.5	10	120	25	3	1.25	0.5	30.65034	
54	Post Nat. I	120	3	1	200	6	11	14	200	25	3	1.33	0.67	37.84939	
55	Product 15	100	3	0	320	1	20	3	45	100	3	1	1	41.50364	
56	Puffed Rice	50	1	0	0	0	13	0	15	0	3	0.5	1	60.75611	
57	Puffed Wh	50	2	0	0	1	10	0	50	0	3	0.5	1	63.00565	
58	Quaker Oa	100	4	1	135	2	14	6	110	25	3	1	0.5	49.51167	
59	Quaker Oa	100	5	2	0	2.7	-1	-1	110	0	1	1	0.67	50.62939	
60	Raisin Bran	120	3	1	210	5	14	12	240	25	2	1.33	0.67	32.2552	
61	Raisin Nut	100	3	2	140	2.5	10.5	8	140	25	3	1	0.5	39.7034	
62	Raisin Squ	90	2	0	0	2	15	6	110	25	3	1	0.5	56.33314	
63	Rice Chex	110	1	0	240	0	23	2	30	25	1	1.13	41.99893		
64	Rice Krisp	110	2	0	290	0	22	3	35	25	1	1	1	40.58016	
65	Shredded 1	80	2	0	0	3	16	0	95	0	1	0.89	1	58.23559	
66	Shredded 1	90	3	0	0	4	19	0	140	0	1	1	0.67	74.47255	
67	Shredded 1	90	3	0	0	3	20	0	120	0	1	1	0.67	72.80179	
68	Smacks	110	2	1	70	1	9	15	40	25	2	1	0.75	31.23005	
69	Special K	110	6	0	230	1	16	3	55	25	1	1	1	53.13132	
70	Strawberry	90	2	1	15	3	15	5	50	25	2	1	1	59.36399	
71	Total Corn	110	2	1	200	0	21	3	35	100	3	1	1	30.63395	
72	Total Rais	140	4	1	190	4	15	14	230	100	3	1.5	1	28.59299	
73	Total Whole	100	3	1	200	3	16	3	110	100	3	1	1	46.85884	
74	Triplex	110	2	1	260	0	21	3	60	25	3	1	0.75	39.10617	
75	Trix	110	1	1	140	0	13	12	25	25	2	1	1	27.7533	
76	Wheat Ch	100	3	1	230	3	17	3	115	25	1	1	0.67	49.78745	
77	Wheaties	100	3	1	200	3	17	3	110	25	1	1	1	51.59219	
78	Wheaties I	110	2	1	290	1	18	3	60	25	1	1	0.75	36.18756	

a) conjunto de dados

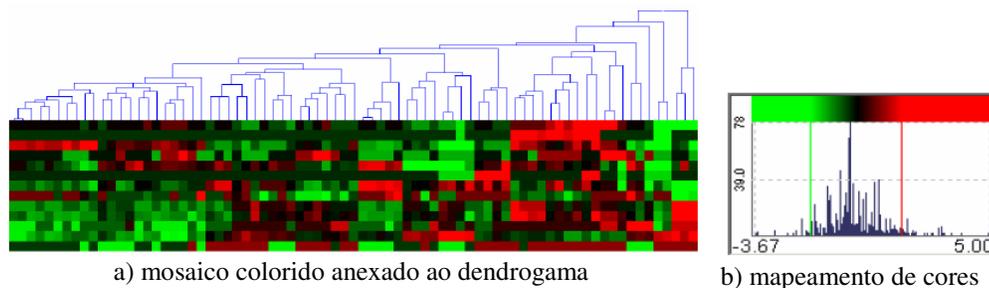


b) mosaico colorido vertical



c) mosaico colorido horizontal

Figura 35 – Mosaico colorido para um conjunto de dados multidimensional.



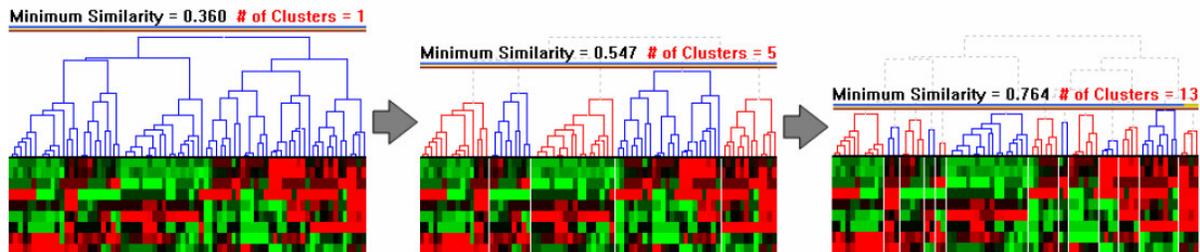
**Figura 36 – Um mosaico colorido anexado a um dendrograma.**

A HCE aglomera os dados e os apresenta por meio de um dendrograma com um mosaico colorido nas suas folhas (Figura 36(a)). O arranjo das linhas e das colunas no mosaico colorido é alterado de acordo com o resultado do algoritmo de aglomeração. O mapeamento das cores é especificado por um controle de cores que usa um histograma para todos os valores numéricos na série de dados (Figura 36 (b)). Por *default*, no HCE, um valor elevado tem uma cor vermelha brilhante e um valor baixo tem verde brilhante. O valor médio tem uma cor preta. A linha vermelha vertical especifica o valor acima do qual todos os valores são mapeados à cor vermelha, a mais brilhante, e a linha verde vertical especifica que o valor está abaixo de todos os valores. Quando um valor começa a chegar mais perto do valor médio entre as linhas verdes e vermelhas, a cor torna-se mais escura.

Os usuários do HCE iniciam o uso da ferramenta executando o algoritmo de aglomeração hierárquico que constrói um dendrograma e um mosaico colorido. Com a barra de similaridade (Figura 37), os usuários podem interativamente ajustar um parâmetro (similaridade mínima) para encontrar o número mais natural dos grupos. Outros controles são possíveis, mas estão fora do escopo deste trabalho e são descritos por Jinwook [62].

Arrastando para baixo a barra de similaridade, cuja coordenada determina o ponto inicial mínimo da similaridade, usuários podem filtrar os elementos menos similares. Desta maneira, os usuários podem, facilmente, encontrar os conjuntos dos elementos que são similares o bastante para estarem no mesmo grupo. Os usuários também podem destacar a posição de um grupo no conjunto de dados original apenas clicando no grupo. A Figura 37 mostra o processo de descoberta dos grupos usando a barra de similaridade. A coordenada vertical da barra

determina o valor mínimo da similaridade. Os usuários podem arrastar para baixo a barra para filtrar os itens que estão distantes de um grupo. Os valores mínimos da similaridade mudaram de 0.36 para 0.764 neste exemplo para separar 1 grupo grande em 13 grupos pequenos.



**Figura 37 – Barra de Similaridade Mínima.**

Apesar de vários pacotes estatísticos de software oferecerem recursos de aglomeração, quase todos implementam somente o algoritmo e produzem o resultado visual estático dos dendrogramas. O HCE, ao contrário, permite o uso dos recursos de aglomeração de forma interativa, permitindo a exploração dinâmica dos grupos de estudos. As características dos estudos podem ser então exportadas para a ferramenta e exploradas para construir o melhor agrupamento de contextos para os resultados em mãos.

Considere o exemplo mostrado na Figura 38, que apresenta como os contextos dos estudos apresentados em 11 artigos foram agrupados automaticamente pela ferramenta. A Seção 4.4.3 descreve em detalhes como este agrupamento reflete o agrupamento feito de forma manual e independente por um grupo de especialistas na área.

É importante notar o quanto o ferramental flexibiliza a análise de contexto. É possível a qualquer momento modificar os pesos dos atributos de contexto de maior interesse, ou simplesmente solicitar que certos atributos sejam desconsiderados. Desta maneira, o algoritmo poderá ser facilmente reprogramado para considerar somente certos aspectos na análise do contexto dos estudos. Isto permite que um cientista possa focar em certos atributos teóricos, experimentais ou humanos de seu interesse. Por exemplo, se aglomerarmos os contextos que consideramos anteriormente usando somente alguns aspectos teóricos dos estudos obteríamos os grupos mostrados na Figura 39.



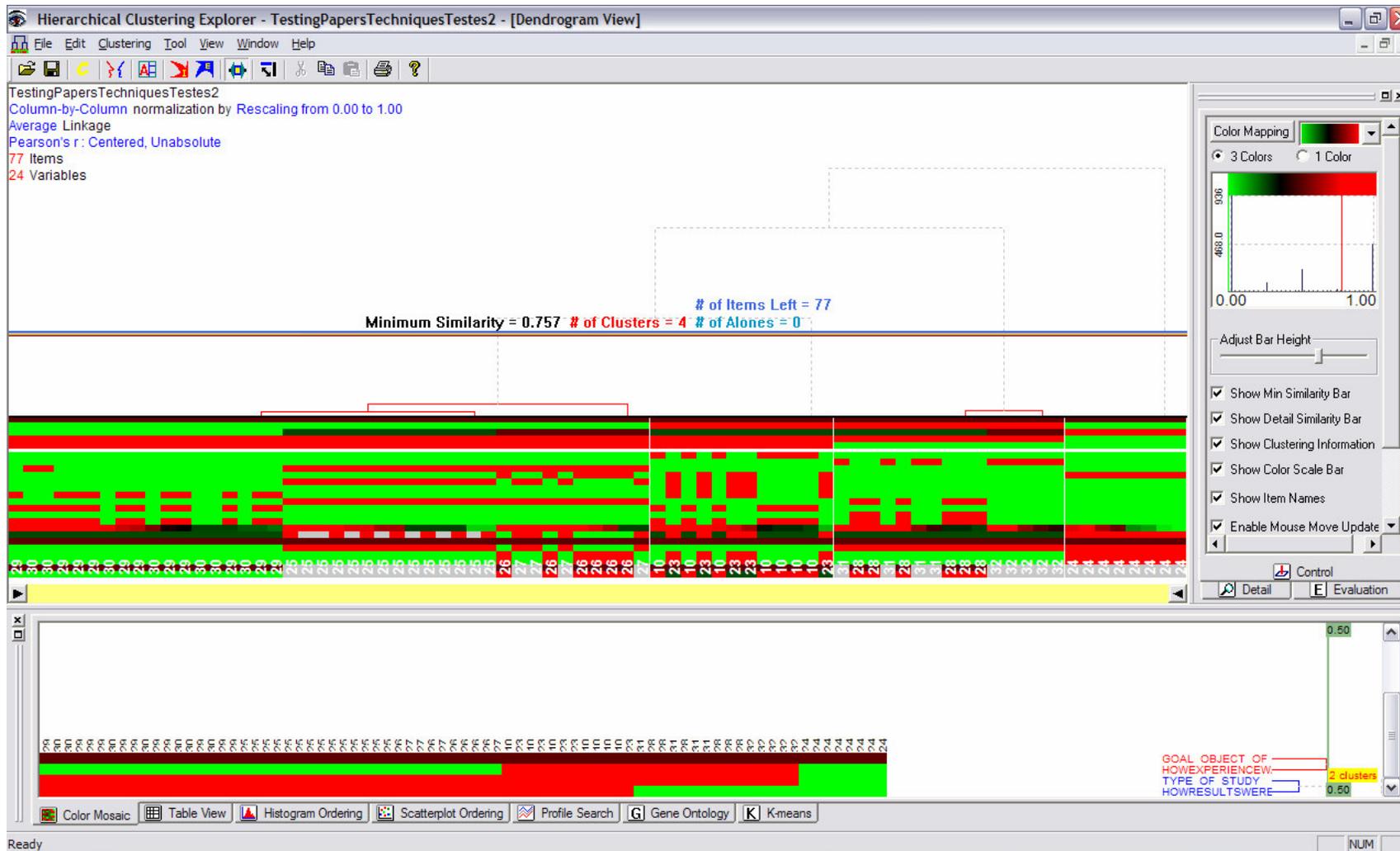


Figura 39 -- Ferramenta HCE –Aglomerando Contextos baseados nos Atributos Teóricos

#### 4.4.3 Exemplo de uso Ferramenta HCE para Exploração da Base de Resultados

Esta seção descreve em um estudo de viabilidade, mostrando o uso das ferramentas de apoio à abordagem SecESE. A metodologia que utilizamos foi:

1. Escolha de um artigo que fizesse uma revisão informal de artigos experimentais;
2. Escolha de alguns dos artigos da revisão;
3. Extração das informações destes artigos;
4. Aplicação do ferramental apresentado anteriormente para analisar estas informações utilizando a abordagem SecESE.

Para o primeiro passo, escolhemos um artigo que faz uma revisão de 25 anos de Experimentos em Técnicas de Teste (ArtigoTTE) [100]. Este artigo analisa o nível da maturidade do conhecimento sobre técnicas de teste examinando estudos experimentais existentes sobre estas técnicas. Eles analisaram os resultados destes artigos, e obtiveram uma classificação do conhecimento sobre técnicas de teste. Como um resultado desta análise, o artigo endereça os seguintes objetivos:

- O objetivo principal é compilar o corpo do conhecimento em técnicas de teste e seu nível da maturidade, de tal maneira que a informação poderia ser útil aos desenvolvedores para identificar as condições da aplicabilidade das diferentes técnicas de teste.
- O segundo objetivo é fornecer um retrato de quais aspectos das técnicas de teste têm sido experimentalmente avaliadas, quais necessitam mais estudos experimentais e que ainda não foram considerados, de tal maneira que os pesquisadores possam usar esta informação para focar sua pesquisa sobre esta questão, aumentando a maturidade do conhecimento na área.

Para o segundo passo, escolhemos 11 artigos dos 24 artigos do ArtigoTTE. Estes artigos foram escolhidos porque tínhamos conhecimento prévio do tópico, isto é, as áreas de teste de que já tínhamos certo conhecimento; neste caso, os artigos relacionados com Técnicas de Testes Estruturais, Funcionais e Teste de Mutação. A lista de artigos escolhida foi a seguinte:

- 
1. Victor R. Basili, Richard Selby, Comparing the Effectiveness of Software Testing Strategies, *IEEE Transactions on Software Engineering*, 13(12):1278-1296, December 1987.
  2. Elaine J. Weyuker. The cost of data flow testing: An empirical study. *IEEE Transactions on Software Engineering*, 16(2): 121–128, February, 1990.
  3. James M. Bieman, Janet L. Schultz. An empirical evaluation (and specification) of the all-du-paths testing criterion. *IEE Software Engineering Journal*,7(1):43–51, January,1992.
  4. Phyllis G. Frankl, Stewart N. Weiss. An experimental comparison of the Effectiveness of branch testing and data flow testing. *IEEE Transactions on Software Engineering*, 19(8): 774–787, August, 1993.
  5. Monica Hutchins , Herb Foster , Tarak Goradia , Thomas Ostrand. Experiments on the Effectiveness of data-flow and control-flow-based test adequacy criteria. *Proceedings of the 16th International Conference on Software Engineering*. 191–200, Sorrento, Italy, May, 1994.
  6. Phyllis G. Frankl, Oleg Iakounenko. Further empirical studies of test Effectiveness. In *Proceedings of the ACM SIGSOFT International Symposium on Foundations on Software Engineering*, 153–162, Lake Buena Vista, Florida, USA, November, 1998.
  7. Glenford J. Myers. A controlled experiment in program testing and code walkthroughs/inspections. *Communications of the ACM*, 21(9): 760–768, September, 1978.
  8. Erik Kamsties, Christopher M. Lott. An empirical evaluation of three defect-detection techniques. *Proceedings of the Fifth European Software Engineering Conference*, pages 362-383, Sitges, Spain, 1995.
  9. Murray Wood, Marc Roper, Andrew Brooks, James Miller. Comparing and combining software defect detection techniques: A replicated empirical study. *ACM SIGSOFT Software Engineering Notes*, 22 (6): 262 – 277, November, 1997
  10. A. Jefferson Offutt, Stephen D. Lee. An empirical evaluation of weak mutation. *IEEE Transactions on Software Engineering*, 20(5): 337–344, May, 1994
  11. A. Jefferson Offutt, Ammei Lee, Gregg Rothermel, Roland H. Untch, Christian Zapf. An experimental determination of sufficient mutant operators. *ACM Transactions on Software Engineering and Methodology*, 5(2): 99–118, April, 1996.

As figuras mostradas na Seção 4.2.2 ilustram a extração de informações desses artigos. Das Figura 19 a Figura 29 mostramos algumas telas da coleta das informações de análise dos artigos. A Figura 19 e a Figura 20 mostram a definição do contexto para a análise. A Figura 21 mostra como coletamos os detalhes gerais dos artigos; nesta tela também coletamos as

informações sobre a qualidade do artigo. A Figura 22 mostra a tela para coletar o que denominamos informações gerais sobre o contexto do estudo no artigo. A Figura 23 mostra a coleta das variáveis dependentes e independentes do estudo. A Figura 24 mostra a coleta dos detalhes da instrumentação. A coleta dos produtos do trabalho é mostrada na Figura 25. Outras informações que o analista necessita coletar podem ser inseridas na base de dados como mostrado na Figura 26; neste caso, coletar a informação de se os defeitos analisados em cada programa foram semeados ou não. Para cada artigo, coletamos além das informações de contexto, os resultados. A coleta dos resultados é mostrada na Figura 27 e a formalização destes resultados é mostrada na Figura 28.

Depois da formalização dos resultados pudemos então executar a análise dos resultados e das descrições do contexto (Figura 29). Os resultados podem ser filtrados e explorados a partir dos objetivos para a análise. Para este estudo de viabilidade não definimos objetivos refinados para a análise porque nosso objetivo era somente avaliar alguns aspectos da abordagem SecESE. Uma análise que fizemos para exploração de resultados em outro contexto é mostrada em um estudo de viabilidade descrito no próximo capítulo.

O analista pode usar o *Treemap* para analisar ao mesmo tempo os resultados e os contextos, e complementar a análise com o HCE que auxilia na exploração da distância e de outras características dos contextos. É possível fazer muitos tipos de combinação de análise; por exemplo, a saída do algoritmo de aglomeração pode fornecer pistas de como explorar os resultados.

Na Figura 33 criamos uma hierarquia na qual o atributo “Como a prática foi aplicada” é o atributo que divide os resultados. Os valores possíveis estão descritos na Tabela 15. Do lado direito da tela da Figura 33, existe uma lista de atributos que podem ser escolhidos para compor a hierarquia; qualquer atributo pode ser escolhido, mas o analista deve estabelecer sobre qual perspectiva da árvore de distância (Figura 16) ele deseja agrupar os contextos.

**Tabela 15- Critérios de Qualidade - Fator Como a prática foi aplicada.**

Critério	Descrição	Valor
Como a Prática foi Aplicada	Não é claro	0
	Um Experimento	3
	Uma série de experimentos	4
	Um projeto piloto	4
	Uma série de projetos pilotos	5
	Um projeto em produção	5
	Uma série de projetos em produção	7

Trabalhando com o contexto usando o algoritmo de aglomeração, validamos o agrupamento pelo agrupamento feito pelas autoras do artigo da revisão. Usamos os seguintes atributos em um arquivo de entrada para realizar a análise dos agrupamentos dos contextos nos 11 artigos analisados:

- Tipo do estudo;
- Descrição do tópico;
- Objeto do estudo;
- Categoria de Participantes (estudantes, profissionais, etc.);
- Modalidade do trabalho dos participantes (indivíduo, equipe, etc.);
- Categoria da tarefa (criar, analisar, planejar, etc.)
- Produtos de trabalho (código, exigências, projeto, etc.);
- Origem do instrumento;
- O nome de cada variável dependente e independente após a verificação de consistência semântica;
- A técnica de teste estudada no artigo (funcional, leitura do código, estrutural, etc.).

Na lista acima, devemos observar duas particularidades. A primeira é que cada estudo envolve diversas variáveis dependentes e independentes. A segunda é que o número e o tipo de variáveis variam de estudo para estudo. A fim de inserir pesos na abordagem SecESE de aglomeração criamos um registro para cada variável de cada estudo em nosso arquivo de

dados. Desta maneira, um estudo que envolve cinco variáveis dependentes e independentes, por exemplo, tem cinco registros no arquivo de dados.

A Figura 38 mostra o agrupamento que foi mostrado pela ferramenta usando os atributos acima. Foi praticamente o mesmo agrupamento feito por Juristo et. al. [100], como especificado na Tabela 16.

**Tabela 16 - Grupos de Estudos definidos por Juristo et. al.**

Estudos	Grupos Ferramenta HCE	Grupos Juristo
Weyuker (23)	Grupo 1	Grupo 1
Bieman and Schultz (24)		
Frankl_Weiss (25)	Grupo 2	Grupo 2
Hutchins (26)		
Frankl_Iak (27)		
<b>Myers (28)</b>	Grupo 4	Grupo 3
Basili (10)	Grupo 3	
Kamties_Lott (29)		
Wood (30)		
Offuf and Lee94 – Mutation (31)	Grupo 4	Grupo 4
Offuf et. al. 96 – Mutation (32)		

Uma questão importante é o peso do atributo da técnica de teste. Este atributo define os tratamentos do estudo; assim, tem-se que considerá-lo como o atributo mais importante para agrupar os estudos. A suposição foi confirmada por entrevistas que conduzimos com os autores do ArtigoTTE. Para isto criamos uma coluna para cada técnica envolvida nos estudos. Desta maneira, estamos reforçando o peso dado aos atributos do contexto categorizados como “instrumentais” na árvore conceitual de distância mostrada na Figura 16. É importante observar que poderíamos também ter solicitado ao algoritmo para considerar outros atributos como mais fortes no processo de aglomeração baseado nos atributos teóricos, experimentais ou humanos do contexto.

---

O arquivo de entrada foi aberto na ferramenta de HCE e executamos o algoritmo de aglomeração hierárquico usando o coeficiente de correlação de Pearson como a medida da distância. A barra de similaridade mínima (MSB) começa sempre com 50%, onde dois grupos principais são mostrados. Um grupo incluiu o Grupo 1 e o outro incluiu os outros grupos 2, 3 e 4 do ArtigoTTE (veja a Tabela 16). A fim de obter mais grupos, movemos o MSB, levantando a medida interna da similaridade dentro dos conjuntos. Obtivemos quatro grupos para um MSB entre 56% e 60%. A Figura 38 mostra como a ferramenta de HCE usa as cores do dendrograma para destacar os grupos.

Analisando a Figura 38, podemos ver que há somente um artigo que foi classificado em uma maneira diferente do que o do ArtigoTTE. O estudo de Myers foi removido do ArtigoTTE porque não continha todas as informações de contexto que elas necessitavam para a análise, e na abordagem SecESE o artigo não foi agrupado no mesmo grupo que Juristo et. al. tinham agrupado; todos os outros estudos foram agrupados corretamente pelo algoritmo. Discutimos esta anomalia com os autores do ArtigoTTE, que disseram que este artigo foi excluído mais tarde de sua análise por causa da falta de alguns detalhes da informação do contexto, especialmente nos detalhes sobre as técnicas estudadas. Este é um resultado bom já que os outros 10 artigos foram classificados corretamente.

Alguns resultados interessantes foram obtidos em valores diferentes de MSB. Para um MSB maior, de 60.5% por exemplo, obtivemos cinco conjuntos. O grupo 1, com os artigos Bieman & Schultz e Weyuker, foi dividido em 2 grupos (ver a Tabela 16). Olhando os dados, percebemos que os dois artigos têm uma técnica em comum, mas o estudo de Weyuker incluiu outras três técnicas que não foram cobertas por Bieman & Schultz.

Em seguida, investigamos outras medidas da distância e obtivemos resultados similares. Acreditamos que o agrupamento dos estudos variará pouco com as métricas de distância usadas (Manhattam, Euclidiana, e Pearson).

Um outro ponto importante a mencionar é que o mosaico da cor é muito útil para visualizar as similaridades e as diferenças entre contextos. Por exemplo, a Figura 38 mostra que este conjunto de estudos é muito uniforme em muitos dos atributos do contexto. O estudo de

viabilidade ilustrou a utilidade de usar a abordagem SecESE para compreender rapidamente como uma grande quantidade de estudos podem ser comparados. Queremos enfatizar a importância de se usar uma ferramenta interativa de visualização para esta tarefa. Como um exemplo, considere uma amostra do arquivo de dados (reduzido no número de linhas e de colunas) mostrado na Figura 40. Pode-se verificar como é difícil aglomerar manualmente os estudos e ver uma relação entre os estudos mesmo quando somente alguns atributos são usados.

Study	Functional	Structural	Code Reading	None	All-c-uses	All-p-uses	All-uses	All-du-paths	All-Edges	Sentence Coverage	Mutation
Basili	X		X							X	
Weyuker					X	X	X	X			
Bieman								X			
FranklWeiss				X			X		X		
Hutchins				X				X	X		
Frankl lak				X			X		X		
Myers	X	X		X							
Kamsties Lott	X		X		X						
Wood	X		X						X		
Offut and Lee											X
Offut et al											X

**Figura 40 - Técnicas estudadas em cada artigo – exemplo de aglomeração Manual**

Apresentamos neste capítulo a evidência de que o uso de uma abordagem sistemática para coletar a informação do contexto combinada com técnicas de aglomeração pode agrupar estudos da mesma maneira que um perito. Esta abordagem cria possibilidades interessantes tais como usar aglomeração interativa para avaliar a generalidade das evidências por meio dos estudos, e para usar um centróide do conjunto identificar o contexto típico para um conjunto de evidências. É importante notar que a abordagem SecESE pode ser usada juntamente com os outros métodos para combinar resultados dos estudos. Pode ser útil, por exemplo, para analisar os dados coletados para revisões sistemáticas.

## Capítulo 5

### Avaliação da SecESE

#### 5.1 Estudo de viabilidade: Artigos de Detecção de Defeitos

O estado atual da prática para abstrair a informação por meio de diversos estudos é executar uma busca na literatura, revendo a literatura relevante e construindo um sumário textual da evidência relacionada a uma questão dada. O conceito de “folclore” tem sido usado como uma maneira de focar tal busca de literatura de forma que seja útil para combinar dados e produzir hipóteses bem formadas e experimentalmente apoiadas [119].

O folclore pode ser definido como as lições aprendidas, informais e subjetivas, baseadas em experiências pessoais. Esta abordagem envolve considerar um folclore como um tipo de hipótese e identificar os artigos que a apóiam ou a rejeitam. No estudo de um conjunto de artigos relacionados, o pesquisador pode construir a evidência a favor ou contra o folclore, refinar o folclore em hipóteses testáveis, ou reconhecer as variáveis do contexto que os diferenciaram, quando e onde as hipóteses eram verdadeiras ou falsas (criando hipóteses mais específicas).

Nesta seção, apresentamos um exemplo de avaliação de folclore como um exemplo de como as hipóteses testáveis podem ser geradas por tal busca de literatura. Escolhemos a área de defeitos de software como tópico já que este é especialmente rico em folclores. Os dois folclores que identificamos foram:

- 1. Existem alguns padrões de classes de defeitos em classes de projetos.**
- 2. A grande maioria dos defeitos são defeitos de interface.**

Os dois folclores são importantes porque juntos indicam a base para muito do que é feito na análise de defeitos. Se houver padrões nos defeitos que encontramos nos projetos, então a análise dos defeitos em um projeto, em um ambiente particular, nos permitirá compreender quais técnicas, métodos e processos aplicar nesse ambiente no futuro. Se a maioria dos defeitos são defeitos de interface, então as técnicas de garantia de qualidade precisam ter o

---

objetivo de impedir ou de identificar defeitos de interface. Para manter a simplicidade do exemplo aplicamos a idéia em três artigos que tínhamos experiência prévia significativa:

- Endres75: um dos primeiros artigos em defeito de software. Descreve a liberação de uma versão de um sistema em operação no qual aproximadamente 500 módulos (140K linhas de código assembler) foram afetados pela modificação. Os defeitos foram classificados como sendo específicos do problema, específicos de implementação, ou específicos do texto. Nessa análise, os defeitos específicos do problema são considerados como devido a interpretação equivocada dos requisitos [5].
- Weiss/Basili85: Este artigo trata de um estudo de três projetos no SEL (Software Engineering Laboratory) na NASA/GSFC. Todos os projetos lidaram com o desenvolvimento de requisitos, projeto e código do software de apoio em terra para controle de naves espaciais não pilotadas, e eram de 50K a 120K linhas de código Fortran. Uma das questões exploradas especificamente no estudo foi a distribuição dos defeitos pela origem do erro, isto é, de acordo com a fase em que o erro ocorreu (requisitos, especificação funcional, projeto ou codificação de componentes múltiplos, de projeto ou de codificação de um único componente), língua, ambiente [39].
- Basili/Pericone84: Este artigo trata do desenvolvimento e a evolução de três anos de um programa de propósito geral para o “Satellite Planning Studies” no SEL. O sistema tinha 90K linhas de código fonte Fortran e os requisitos mudaram e evoluíram durante este tempo. A mesma questão do estudo Weiss/Basili85 foi estudada neste trabalho, mas as fases foram um pouco diferentes (requisitos, especificação funcional, projeto, codificação e evolução) [129].

O estudo Endres75 teve somente um projeto; as maiores fontes de defeitos (46%) que o projeto teve foram associadas às interpretações equivocadas do domínio do problema. No estudo Weiss/Basili85, os três projetos tiveram perfis muito similares, a maioria dos defeitos (entre 56% e 72%) ocorreu durante projeto e codificação de um único componente. No

---

estudo Basili/Pericone84, embora os dados tenham sido recolhidos no mesmo ambiente do SEL, a maioria dos defeitos foi causada por interpretações equivocadas dos requisitos (55%).

A diferença nos resultados entre estes estudos nos permite examinar o impacto das variáveis de contexto. Em Weiss/Basili85, a organização tinha previamente desenvolvido muitos projetos similares, de modo que, embora os requisitos variassem, a organização tinha experiência em desenvolver esse tipo de sistema. Assim, o fato que a maioria dos defeitos estava ocorrendo nas fases de codificação era devido ao fato que os requisitos e projeto de alto nível eram relativamente bem compreendidos pelos desenvolvedores, mas, os novos desenvolvedores da organização é que foram usados para desenvolver o código. Por outro lado, os estudos Endres75 e Basili/Pericone84 envolveram projetos com requisitos relativamente novos e menor compreensão dos requisitos. Isto permite que postulemos duas hipóteses novas:

**1a) Em projetos novos, a maior fonte de defeitos é pouco entendimento dos requisitos.** [Apoiada por: Endres75, Basili/Pericone84; Contradita por: Nenhum].

**1b) Em projetos que organização já possui experiência na aplicação, a maior fonte de defeitos é a codificação** [Apoiada por: Weiss/Basili85; Contradita por: Nenhum].

Esta análise também permite o reconhecimento de que as hipóteses como originalmente descritas são também apoiadas. No que diz respeito à segunda hipótese, a questão é a distribuição de defeitos de interface. O estudo Endres75 relatou que somente 15% dos defeitos encontrados eram defeitos de interface. Entretanto, ele definiu os defeitos de interface baseado no número de componentes modificados, isto é, um defeito é um defeito de interface se mais de um componente (módulo) deve ser modificado ao se consertar o problema. Podemos pensar como uma “implementação de um defeito de interface”. Usando a mesma definição, o estudo Basili/Pericone84 encontrou que somente 11% dos defeitos eram “implementação de defeitos de interface”. Entretanto, no estudo Basili/Pericone84 os autores usaram uma segunda definição para defeito de interface, associada com o número de componentes examinados; isto é, é um defeito de interface se mais de um componente (módulo) tiver que ser examinado a fim de projetar a mudança. Podemos chamar este tipo de

“um defeito de interface de projeto”. Esta é uma definição muito mais inclusiva e a maioria de definições encontrar-se-ia em algum lugar entre estas duas definições. Usando esta última definição, no estudo Basili/Pericone84 os autores encontraram que 39% dos defeitos eram “defeitos de interface de projeto”. Assim, em um ou outro caso, o folclore não é verdadeiro, significando que uma hipótese útil que descreve o comportamento do defeito deve ser formulada considerando exatamente a condição oposta:

## **2'. Defeitos de Interface não representam a maior parte dos defeitos em um sistema.**

[Apoiada por: Endres75, Basili/Pericone84; Contradita por: Nenhum].

Esses exemplos mostram o benefício de acumular resultados dos estudos múltiplos em múltiplos artigos, reconhecendo como as variáveis de contexto podem fornecer introspecções para o refinamento das hipóteses, assim como criar novas hipóteses, mas:

- A falta de formalidade no processo permite que o analista perca alguns detalhes mais facilmente. Na análise acima, a integração dos estudos foi feita por alguém com um alto grau de conhecimento do contexto a respeito de todos os estudos descritos, mas este não é sempre o caso.
- Como pode ser observado na descrição dos contextos, é difícil abstrair as mesmas informações de contexto de todos os artigos quando não existe um conjunto básico de informações a serem buscadas e extraídas do texto.
- A análise não é reusável; as revisões da literatura são geralmente feitas para endereçar uma questão particular e, se uma questão relacionada, porém diferente, for do interesse de um outro pesquisador, a análise deve ser refeita.

### **5.1.1 Detecção de Defeitos – Uma análise mais formal**

Para permitir uma comparação entre o uso da SecESE, e o exemplo dado da análise baseada em folclore, descrita acima, selecionamos novamente defeitos de software.

Primeiramente, usamos um conjunto relativamente pequeno de artigos no qual tínhamos alguma experiência sobre o contexto da aplicação, para investigar se os resultados que poderiam ser obtidos da aplicação da SecESE seriam comparáveis a uma opinião de um

---

perito. Estes resultados são descritos nas seções seguintes. Depois, tornamos a fazer uma análise com um conjunto maior de artigos e de autores. Na terceira análise, comparamos os resultados da SecESE com estes artigos e os resultados de um perito externo (que fez uma revisão sobre alguns artigos sobre defeitos); nosso objetivo era avaliar a completude e a repetibilidade da abordagem SecESE.

### 5.1.2 Resultados da Análise

Para permitir a comparação com a abordagem baseada em folclore descrita anteriormente, aplicamos a SecESE usando somente os três estudos da Seção 5.1: Endres75, Basili/Pericone84, e Weiss/Basili85. A abordagem foi aplicada por um pesquisador que não tinha experiência direta com os artigos.

Setenta e sete resultados foram extraídos destes artigos. Durante a análise, usamos somente os resultados que foram apoiados por dados experimentais e filtramos a análise para incluir somente as relações que foram apoiadas por mais de um artigo. A ferramenta *Treemap*<sup>4</sup> foi usada para explorar os resultados.

No *Treemap*, a primeira hierarquia criada foi baseada no atributo “variável dependente”. Encontramos três grupos grandes de resultados relacionados aos defeitos, às mudanças, e ao esforço. Analisando em detalhe a variável dependente “defeitos”, criamos um nível novo da hierarquia baseado no atributo “variável independente”. Analisando a Figura 41, podemos ver as principais variáveis independentes relacionadas aos defeitos: interface, tamanho, esforço, interpretação equivocada da especificação, modificações. Na Figura 41, cada retângulo pequeno representa um resultado individual, com o texto mostrado dentro do bloco. As cores identificam os artigos dos quais estes resultados foram extraídos. Os agrupamentos maiores aglomeram todos os resultados sobre uma variável independente em comum.

Nesta fase da análise, temos uma lista sistemática de relações potenciais no foco do interesse (defeitos de software) que pode ou não representar uma informação nova para o pesquisador. Cada relacionamento pode ser explorado mais detalhadamente para examinar o

contexto no qual o relacionamento foi estabelecido e o nível de suporte a estes resultados. Ao escolher analisar a “interface” como variável independente, obtivemos os resultados mostrados na Figura 42. Podemos ver que dois dos três artigos não apóiam a hipótese que a maioria de defeitos é defeitos de interface.

Para tentar resolver esta discrepância, analisamos mais detalhadamente as definições de “interface” em cada artigo. Nos dois artigos que não apóiam a hipótese, a definição de defeitos de interface é baseada no número de componentes modificados, isto é, um defeito é um defeito de interface se mais de um componente (módulo) deve ser modificado no conserto de um problema. No terceiro artigo, os defeitos de interface estavam associados com as estruturas que existem fora do módulo. Podemos dizer então que, usando a primeira definição, os defeitos de interface não são a maioria dos defeitos neste estudo. Explorando as outras variáveis independentes, podemos estabelecer as seguintes conclusões adicionais:

- A. Os defeitos de interface não são a maioria dos defeitos se usarmos a definição de defeitos de interface que é baseada no número dos componentes modificados [Apoiada por: Endres75, Basili/Pericone84, e Weiss/Basili85; Contradita por: Nenhum].
- B. As interpretações equivocadas das especificações compõem a maioria dos defeitos quando os desenvolvedores não são experientes no domínio da aplicação. [Apoiada por: Endres75 e Basili/Pericone84; Contradita por: Weiss/Basili85].
- C. Tamanho, somente, não é um fator para determinar a predisposição à ocorrência de defeitos. [Apoiada por: Endres75, Basili/Pericone84, e Weiss/Basili85; Contradita por: Nenhum].

Comparando estas hipóteses aos resultados da abordagem baseada em folclore, percebemos que a hipótese A, acima, confirma a hipótese refinada 2' na seção anterior. A hipótese B, acima, também mapeia diretamente à hipótese 1a na seção anterior. A hipótese C, acima, é um exemplo concreto do caso geral descrito na hipótese 1 (a saber, que existem padrões consistentes que podem ser encontrados para descrever ocorrências do defeito de software).

---

<sup>4</sup> <http://www.cs.umd.edu/hcil/Treemap/>

Defect						
Defect			Changes		Interface	
The number of errors that occurs per module is, on the average, very small	Modified modules had a high percentage of errors of commission and a small percentage of errors of omission with a higher percentage of data and initialization errors	The most common type of error is one made in the design or implementation of a single component of the system.	Errors Corrections are more frequent than modifications	89 Percent of errors can be corrected by changing only one module	Interfaces appear to be the major source of errors regardless of the module type.	Interfaces are not a major source of errors
Modified and new modules behave similarly except for the type of errors prevalent in each and the amount of effort required to correct an error	More than 85% of the errors could be corrected by changing only one module per error	Modified and New Modules have similar number of Errors.	The greater the number of requirements changes does not imply greater number of requirements errors	Relatively few changes result in errors	Interfaces and requirements understandings errors are more difficult to correct than others	39% of the errors are Interface Errors
New modules have and equal number of errors of omission and commission and a higher percentage of control errors	For each type of error there exist different causes and therefore different measures have to be taken to prevent them. There is no single cure at all.		Despite a significant number of requirements changes imposed on some projects, there is no corresponding increase in frequency of requirements	85% of the errors could be corrected by changing only one module per error. It contradicts the preconception we have of the interdependence of the		
Misunderstandings of Specifications			Effort		Size	
The majority of errors are the result of functional specification (incorrect or misinterpreted)	Modified modules appear to be more susceptible to errors due to the misunderstanding of the specifications	Software development environment, the programming language used, misunderstandings of interface, project size, and misunderstandings of specifications do not	50% of the total effort required for error correction occurred in modified modules	Errors contained in modified modules requires more effort to correct than those in new modules, although the two classes contained approximately the same number of	Larger modules have more errors	There is a higher error rate in smaller sized modules
The majority of errors are the result of functional specification (incorrect or misinterpreted) and within this the majority of the errors involved modified modules	Misunderstanding of a module's specifications or requirements constituted the majority of detected errors.	36% of Errors have the source on Functional Specification Incorrectness or Misinterpreted. 24% of these erroes occurred on Modified Modules and 12	Errors ocurring in new modules required less effort to correct than those in modified modules	SEL programmers tend to spend their time finding and correcting many small errors made while designing or implementing single routines, rather than struggling with a few	Software development environment, the programming language used, misunderstandings of interface, project size, and misunderstandings of	

Figura 41 – Conjunto de Resultados – Hierarquia Primeiro Nível: Variável Dependente e Segundo Nível Variável Independente

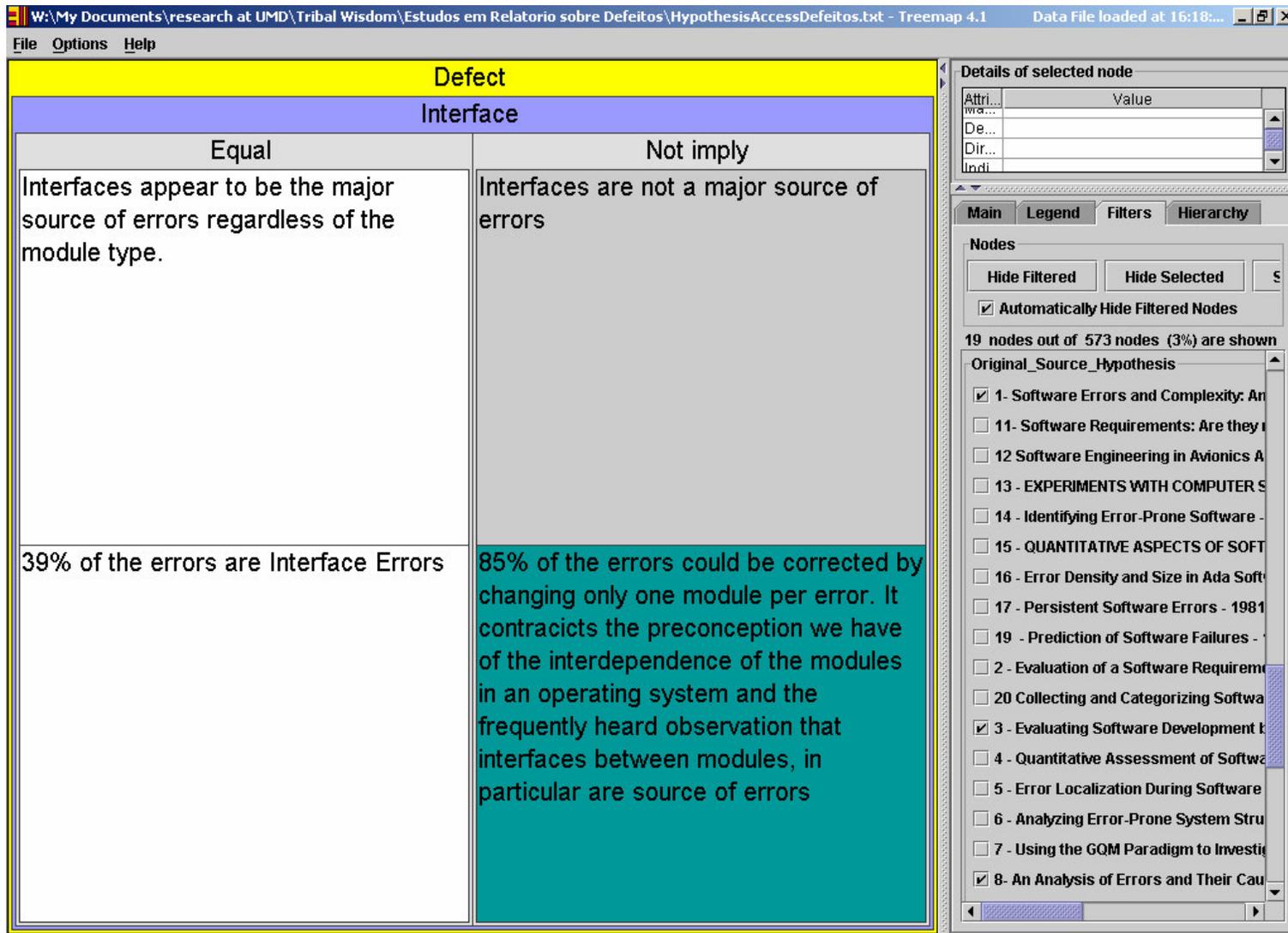


Figura 42 – Resultados que mostram a relação entre Defeitos de Interface e Defeitos de Software.

---

### 5.1.3 Uma Base de Resultados Expandida

Uma das vantagens da abordagem mais formal é que a base estruturada de resultados é reusável e expansível na medida da necessidade do pesquisador. Podemos facilmente fazer uma nova análise somente adicionando artigos na base. Especificamente, adicionamos dados dos seguintes artigos:

- 1- Rombach/Basili87: Este estudo foi conduzido em um ambiente de manutenção de uma grande empresa para sistemas comerciais [40].
- 2- Selby/Basili88: Única versão de uma ferramenta de “*code library*” [112].
- 3- Selby/Basili91: Única versão de uma ferramenta de “*code library*” [113].
- 4- Mashiko/Basili97: Um conjunto de quatro projetos lidando com software de comunicação [138].

Esta adição de artigos trouxe novos resultados para a base, que agora passou a ser de 223 resultados. Novamente, estes resultados foram filtrados para este exemplo, removendo as indicações de opinião que não tiveram dados experimentais validando-as, limitando a análise àqueles resultados que foram endereçados por mais de um artigo.

O processo da análise ajuda a iniciar a compreensão do espaço da solução coberto por estes estudos. Por exemplo, enquanto Mashiko/Basili97 tem alguns resultados sobre interfaces e defeitos, ao resolver os problemas de inconsistências semânticas torna-se aparente que todas estão relacionadas às interfaces externas (isto é, defeitos na interface entre o produto e seu sistema externo) ou interface humano-computador. Estes fenômenos são diferentes dos cobertos nos outros três artigos.

A análise dos artigos restantes não adicionou nenhum resultado novo à lista sobre interpretações equivocadas na especificação, mas aumentou o nível de confiança na hipótese relacionada ao tamanho, fornecendo hipóteses a partir de novos contextos adicionais que corroboram as conclusões iniciais.

#### 5.1.4 Inclusão de artigos de outros autores

Para mostrar que é possível se refazer uma análise feita anteriormente por outro autor, decidimos replicar o estudo feito por Marick [21], no qual diversas bases de dados sobre defeitos foram revisadas a fim de formular conclusões sobre questões comuns a comportamentos do defeito. Especificamente, adicionamos à nossa base as informações vindas dos seguintes artigos, relacionados aos artigos tratados no artigo de Marick [21]<sup>5</sup>:

- Basili81: Avaliação do documento de requisitos do A-7 no re-desenvolvimento do software de vôo A-7[130].
- Bell76 contém uma análise de dois contextos. O primeiro era um projeto em uma turma de engenharia de software na pós-graduação. Os estudantes escreveram um conjunto de requisitos de software e um projeto preliminar do sistema para um Student Employment Information System (SEIS). O segundo caso era maior do que o primeiro contexto (1 milhão de instruções de máquina), um sistema de tempo-real de um sistema Ballistic Missile Defense (BMD) [122].
- Dniestrowski78: Os dados relatados neste artigo são sobre o controle de vôo digital do FALCON 20. O código fonte consistiu de: 800 declarações das variáveis, 1.715 declarações de constantes, 3.615 instruções assembler, 9.100 linhas de documentação [1].
- Glass81: Sistemas de programas operacionais para o uso de aviões militares foram analisados neste artigo. O projeto A envolveu 150 programadores no período de maior carga, que resultou em meio milhão de instruções. O projeto B envolveu 30 programadores e aproximadamente 100.000 instruções [115].
- Lipow79: são apresentados os dados de três projetos grandes de software. Foram analisados o efeito em defeitos de determinadas ferramentas preventivas e detectivas e de algumas técnicas [99].

---

<sup>5</sup> É importante notar que estes artigos foram selecionados para fazermos a mesma análise feita por Marick e não para produzir um corpo detalhado e atual sobre o conhecimento do defeito

- 
- Ostrand84: Neste artigo o autor relatou os resultados obtidos da coleta de dados de defeitos por nove meses em um projeto de um editor. O fonte do editor foi de aproximadamente 10.000 linhas de linguagem de alto-nível e também Assembly. Depois da especificação inicial, o projeto e a codificação do programa foram feitos por três programadores em dez meses. A implementação representa aproximadamente 18 pessoas-mês de esforço [124].
  - Potier82: Os programas usados nos experimentos pertencem a uma família de compiladores para a linguagem em tempo real LTR desenvolvida para uma larga escala de máquinas. O número de linhas total é 64.939, com 1.402 defeitos encontrados [31].
  - Rubey75: Os dados apresentados neste artigo foram obtidos a partir de esforços de validação, envolvendo geralmente os programas de controle em tempo real, relativamente pequenos (em média 32K de instruções em linguagem de máquina) [111].
  - Shen85: Este estudo foi baseado na análise detalhada de três produtos desenvolvidos no laboratório de Santa Teresa da IBM. Os produtos estudados foram desenvolvidos e liberados desde 1980. O produto A é uma ferramenta de contagem de métricas de software escrita principalmente em Pascal, o tamanho total é 7K TSI (número total de milhões de instruções, excluindo comentários). O produto B é um compilador escrito em PL/S (um derivado do PL/1), com 94K TSI. O produto C é um sistema de base de dados escrito linguagem Assembly, o tamanho total é 326K TSI [136].
  - Withrow90: O objeto deste estudo foi um software Ada para o comando e o controle de um sistema de comunicações militares. O projeto é composto de 362 pacotes Ada, contendo 114.000 linhas de código. A média do pacote era de 316 linhas de código por módulo [24].

Note que a descrição dos contextos acima mencionados ainda não continha uma formalização que estabelecesse um formato padrão de coleta das mesmas informações de cada contexto, fato que dificultou de certa forma a análise dos resultados vindos destes artigos. A base passou a conter 18 artigos, representando o trabalho de 26 autores; destes

artigos, 396 resultados foram coletados e formalizados. Durante a etapa da análise, a visualização e a mineração de dados foram aplicadas nos resultados formalizados. Oitenta resultados estavam relacionados aos tópicos destacados por Marick como pontos comuns encontrados em sua análise. Os resultados restantes de texto trataram de vários outros tópicos, como manutenibilidade, esforço, confiabilidade, restrições de memória, reusabilidade, coesão e acoplamento, que estão fora do escopo da análise atual. Para manter a comparabilidade, escolhemos analisar somente a interseção de nossa base de resultados com os tópicos identificados por Marick.

Encontramos quase os mesmos resultados que Marick encontrou; ele usou uma abordagem manual para abstrair informações baseadas em evidências experimentais. A Tabela 17 mostra a comparação dos resultados produzidos pela SecESE e pela abordagem manual feita por Marick.

Nesta comparação, podemos ver que a SecESE teve a vantagem de mostrar as ligações entre os dados subjacentes e as conclusões mais explícitas (por exemplo, conclusão número 7). Em outros casos, as conclusões foram mais robustas porque mais fontes e hipóteses mais específicas que resultaram delas poderiam explicitamente ser associadas (por exemplo, as conclusões 1 e 2).

Com a comparação com a abordagem menos formal, encontramos uma falha na SecESE; não estávamos extraindo a informação das figuras. Isto é, as informações que apareciam codificadas como figuras ou gráficos não estavam sendo inseridas na base; sendo assim, algumas conclusões de Marick não apareceram em nossas conclusões porque elas eram baseadas em figuras. Assim, decidimos introduzir esta etapa na extração da informação dos artigos. Inserimos estas informações na base, re-analisamos a base e os resultados finais são mostrados na Tabela 17.

Percebemos neste estudo que podemos replicar resultados de outros relatórios manuais, mas com uma vantagem de que a revisão foi conduzida de acordo com uma abordagem mais explícita.

**Tabela 17 – Comparação dos Resultados produzidos através da SecESE e da abordagem manual por Marick**

No.	Conclusão	Suporte na análise de Marick.	Suporte encontrado na base.
1	Defeitos na lógica de programação (seleção de caminhos) são comuns.	[Dniestrowski78], [Potier82], [Shen85], [Glass81], [Lipow79]	Mesmos que Marick mais: [Basili84], [Ostrand84]. (11 resultados)
2	Defeitos de omissão são importantes.	[Basili84], [Glass81], [Ostrand84].	Mesmos que Marick mais: [Bell76]. (Também [Mashiko97], [Selby91] que não estavam no conjunto de Marick). (29 resultados)
3	Manipulação de dados é mais suscetível a defeitos do que a computação de dados.	[Basili84], [Potier82], [Shen85], [Glass81], [Lipow79], [Ostrand84].	Mesmos que Marick: (8 resultados)
4	Módulos modificados são mais suscetíveis a erros do que módulos novos.	[Basili84] e fontes adicionais.	Suporte de: [Basili84], [Shen85], [Selby91]. (3 resultados)
5	Existe a evidência de que tanto módulos pequenos quanto grandes são mais suscetíveis a erros do que módulos de tamanho mediano.	[Basili84], [Shen85], [Withrow90]	Mesmos que Marick mais [Endres75], [Potier82]. (Também [Selby91], [Selby88] que não estavam no conjunto de Marick). (10 resultados)
6	Em descrições abstratas (requisitos/ especificações funcionais), erros de incorreção, omissões e inconsistências, são mais freqüentes, nesta ordem.	[Basili81], [Bell76].	Mesmos que Marick mais [Basili84], [Weiss85], [Shen85]. (13 resultados)
7	Consertar defeitos causa um número pequeno de novos defeitos.	Não é claro.	[Basili84], [Weiss85], [Potier82], [Ostrand84]. (6 resultados)

## 5.2 Avaliação Experimental em Artigos Gerais

No estudo de viabilidade descrito anteriormente, aplicamos a técnica de extração de informações de artigos, mas, existiam ainda algumas questões abertas sobre esta experiência. Para a investigação destas questões decidimos projetar um experimento; o objetivo em criar o experimento foi:

- Analisar a **técnica de extração da informação** para compreendê-la com respeito à **completude e à consistência** em extrair **resultados e informações de contexto** de artigos sobre estudos experimentais do ponto de vista dos pesquisadores em Engenharia de Software.

As perguntas relacionadas a este objetivo eram:

- Outras pessoas encontram os mesmos resultados nos artigos que o projetista da técnica encontra?
- Outras pessoas classificam os artigos da mesma forma que o projetista da técnica os classifica?
- O esquema de classificação é razoável?
  - Completude: O esquema inclui as características mais importantes do estudo (as que são essenciais para compreensão do contexto do estudo e também dos resultados).
  - Consistência: cada artigo é categorizado usando sempre os mesmos valores para os atributos.
- Qual é o tempo médio gasto em cada artigo?
- A experiência do leitor afeta os resultados?

### 5.2.1 Experimentadores

Pesquisadores na Universidade de Maryland que tinham criado a técnica e eram conseqüentemente peritos em seu uso. Estes pesquisadores executaram tanto este

---

experimento quanto as repetições do experimento descrito nas próximas seções. O projetista da técnica é o oráculo para as respostas.

## 5.2.2 Participantes

Os participantes eram estudantes de pós-graduação da Universidade de Maryland matriculados em uma turma de Engenharia de Software Experimental em 2006. Neste estudo, havia oito participantes. Como será descrito na seção dos procedimentos, cada participante extraiu informações de três artigos.

Coletamos as informações descritas na Tabela 18 de cada um dos participantes. A escala usada para nível de experiência foi a seguinte:

**0 = Nenhuma experiência**

**1 = Experiência somente em sala de aula**

**Deve-se indicar se:**

**a) Aprendeu somente os conceitos em classe;**

**b) Usou o conceito em um trabalho de classe;**

**c) Usou o conceito em um projeto.**

**2 =Experiência Profissional**

**Indique o número de projetos nos quais você executou a atividade.**

**Tabela 18 - Experiência dos Participantes - Experimento 1**

<b>Participantes</b>	<b>Número de anos de Experiência em Engenharia de Software</b>	<b>Experiência em Participar de um Experimento</b>	<b>Experiência em Replicar Experimentos</b>	<b>Experiência Projetando Experimentos</b>
S1	3	0	0	0
S2	25	0	0	0
S3	20	1 a.b	0	0
S4	2	2 (2)	1B	0
S5	4	0	0	0
S6	0	0	1B	0
S8	2	2(1)	0	2(4)

## 5.2.3 Materiais

O material usado no experimento foi o seguinte:

1. Explicação da tarefa: uma explicação breve das etapas que os estudantes devem seguir para participar do experimento;
2. Explicação da técnica: explicação sobre a técnica de extração da informação e o esquema de classificação das informações de contexto, um documento com algumas diretrizes sobre o uso da técnica;
3. Formulários de Coleta: um arquivo do Excel para coleta das informações dos resultados e das informações de contextos.
4. Questionário de Experiência: Um questionário que coleta as informações básicas sobre a experiência do participante;
5. Artigos a serem usados no experimento:
  - a. Jeffrey Parsons, Chad Saunders. Cognitive heuristics in software engineering applying and extending anchoring and adjustment to artifact reuse. *IEEE Transactions on Software Engineering*, 30(12): 873-888, December, 2004.
  - b. Ioannis Samoladas, Ioannis Stamelos, Lefteris Angelis, Apostolos Oikonomou. Open source software development should strive for even greater code maintainability. *Communications of the ACM*, 47(10):83-87, October 2004.
  - c. Marek Vokac, Defect Frequency and Design Patterns: An Empirical Study of Industrial Code. *IEEE Transactions on Software Engineering*, 30(12):904-917, December, 2004.
  - d. Matthias Müller, Two controlled experiments concerning the comparison of pair programming to peer reviews, *Journal of Systems and Software*, 78:166-179, January, 2005.
  - e. A. Gunes Koru, Jeff (Jianhui) Tian. Comparing High-Change Modules and Modules with the Highest Measurement Values in Two Large-Scale Open-Source Products. *IEEE Transactions on Software Engineering*, 31(8): 625-642, August, 2005.
  - f. Richard W. Selby. Enabling Reuse-Based Software Development of Large-Scale Systems. *IEEE Transactions on Software Engineering*, 31(6):495-510, June, 2005.
  - g. A. Pretschner, W. Prenninger, S. Wagner, C. Kühnel, M. Baumgartner, B. Sostawa, R. Zölch, T. Stauner. One Evaluation of Model-Based Testing and its Automation *Proc. 27th Intl. Conf. on Software Engineering (ICSE'05)*, pages 392-401, St. Louis, May, 2005.

- 
- h. Ranjith Purushothaman, Dewayne E. Perry, Toward Understanding the Rhetoric of Small Source Code Changes. *IEEE Transactions on Software Engineering*, 31(6):511-526, June, 2005.
  - i. Kjetil Moløkken-Østvold, Magne Jørgensen. Group Processes in Software Effort Estimation. *Empirical Software Engineering Journal*, 9(4):315-334, 2004.
  - j. Kjetil Molokken-Ostvold, Magne Jorgensen, A Comparison of Software Project Overruns-Flexible versus Sequential Development Models. *IEEE Transactions on Software Engineering*, 31(9): 754-766, September, 2005.
  - k. Tore Dyba. An Empirical Investigation of the Key Factors for Success in Software Process Improvement. *IEEE Transactions on Software Engineering*, 31(5): 410-424,May, 2005.
  - l. Hakan Erdogmus, Maurizio Morisio, Marco Torchiano. On the Effectiveness of the Test-First Approach to Programming. *IEEE Transactions on Software Engineering*, 31(3): 226-237, March, 2005.
  - m. Parastoo Mohagheghi, Bente Anda, Reidar Conradi. Effort estimation of use cases for incremental large-scale software development. *ICSE 2005*, pages 303-311, 2005.
  - n. Michael Yanguo Liu, Issa Traore. Measurement Framework for Software Privilege Protection based on User Interaction Analysis. *11th International Software Metrics Symposium(Metrics 2005)*, Como, Italy, September 2005.
  - o. Trung T. Dinh-Trong, James M. Bieman. The FreeBSD Project: A Replication Case Study of Open Source Development. *IEEE Transactions on Software Engineering*, 31(6): 481-494, June, 2005.
  - p. Martin P. Robillard, Wesley Coelho, Gail C. Murphy. How Effective Developers Investigate Source Code: An Exploratory Study. *IEEE Transactions on Software Engineering*, 30(12): 889-903, December, 2004.
6. Formulário de Consentimento;
  7. Questionário de Pós-Estudo;
  8. Não havia nenhum treinamento formal, mas uma palestra em sala de aula e um artigo sobre a aplicação da técnica usado como exemplo e referência [44].

### 5.2.4 Procedimento

Um questionário de experiência foi distribuído aos participantes para coletar a informação relevante sobre suas experiências do desenvolvimento do software. Uma cópia deste questionário pode ser encontrada no Apêndice A.

O experimento foi parte de uma disciplina e o primeiro trabalho desta disciplina teve como objetivo criar um conjunto de referências anotadas para engenharia de software experimental (experimentos, estudos de caso, pesquisas de opinião, entrevistas, modelos experimentais, métricas, etc.). Dois artigos foram atribuídos a cada estudante para ler, revisar e escrever um resumo anotado do artigo.

Este experimento foi uma continuação ao projeto anotado da bibliografia. A cada um dos estudantes foi pedido que relessem os mesmos dois artigos da primeira parte, assim como um artigo novo (escolhido pelos experimentadores), três no total de artigos a serem analisados.

O projeto experimental foi estabelecido como mostrado na Tabela 19, onde:

- 1) Artigos 1, 8, 10 e 14 foram lidos por 3 dos participantes, todos os outros artigos foram lidos por somente uma pessoa.
- 2) Não controlamos a ordem na qual cada participante deveria analisar os seus 3 artigos; não pedimos que fizessem em alguma ordem específica.
- 3) Foram usados 16 artigos de diferentes tipos (experimentos, pesquisas de opinião, e estudos de caso) dos diferentes tópicos em engenharia de software (reuso, código-aberto, programação em pares, etc.).

**Tabela 19 - Projeto experimental - Experimento 1**

<b>Participantes</b>	1	2	3	4	5	6	7	8
<b>Artigos da Parte 1</b>	1,2	3,4	5,6	7,8	9,10	11,12	13,14	15,16
<b>Novos Artigos</b>	10	10	14	14	1	1	8	8

---

Os estudantes tiveram uma semana para executar as tarefas em casa. Após ter realizado a extração das informações dos artigos, cada participante respondeu um questionário de pós-estudo com perguntas sobre suas impressões sobre a técnica e sugestões sobre como podemos melhorar a técnica e o experimento. Um estudante desistiu do experimento. Assim, tivemos dados de sete participantes com um total de 19 submissões válidas.

Finalmente, a discussão dos resultados foi realizada depois que os participantes completaram a análise e todos os questionários. O objetivo da discussão na sala de aula era mostrar aos participantes o projeto experimental assim como os objetivos e as hipóteses do experimento. Uma análise preliminar dos dados foi também mostrada e discutida com os participantes. Os participantes tiveram a oportunidade de discutir os resultados com os pesquisadores e de esclarecer pontos nebulosos nos dados; esta discussão forneceu uma verificação inicial da validade dos resultados.

### **5.2.5 Coleta de Dados**

Coletamos dados quantitativos e qualitativos. Os dados quantitativos incluíram o tempo requerido para executar a extração usando a técnica e o número de resultados encontrados por um participante e a taxa de acerto na coleta das informações de contexto. Os dados qualitativos foram coletados usando um questionário do pós-estudo e uma sessão de discussão feita com os estudantes. Dados de um oráculo também foram coletados neste experimento para a comparação dos resultados dos participantes. O oráculo neste caso foi o projetista da técnica.

### **5.2.6 Resultados/Lições Aprendidas.**

Os resultados apresentados são baseados em três questões principais:

- 1) Os participantes encontrarão os mesmos resultados nos artigos que o oráculo?
- 2) Os participantes classificarão os artigos da mesma forma que o oráculo?

3) Qual é o tempo médio gasto em cada artigo?

5.2.6.1 Os participantes encontrarão os mesmos resultados nos artigos que o oráculo?

Para responder a esta pergunta, comparamos cada resultado que o oráculo encontrou em cada artigo com os resultados encontrados por cada participante, e dividimos os resultados em três níveis:

- 1) Os textos dos resultados são consistentes (o texto do projetista é similar ao texto do oráculo);
- 2) Nível de Suporte do resultado é consistente.
- 3) Confiança no resultado é consistente.

Além dos níveis da análise, percebemos que havia alguns resultados que foram descritos implicitamente no texto do artigo. Outros foram descritos explicitamente, e então, eram mais fáceis de encontrar e reconhecer no texto, nomeados aqui como resultados explícitos; em qualquer outro caso eles são chamados resultados implícitos.

A Tabela 20 descreve as nossas expectativas e os resultados que obtivemos em cada nível da análise (Texto, Nível de Suporte e Confiança).

A quantidade de resultados encontrados pelos participantes foi baixa para nossas expectativas, Os participantes encontraram em média: **53% dos resultados do artigo e 30% dos “resultados de outros artigos”**. A maioria dos resultados encontrados eram explícitos. A maioria dos resultados implícitos não foi encontrados pelos participantes. Os participantes encontraram **74% dos resultados explícitos e 11% dos resultados implícitos**. Questionamos esse resultado na apresentação do relatório da análise (feita uma semana depois que os participantes submeteram suas respostas), e percebemos que não haviam compreendido realmente a definição de resultados. Os participantes nesta discussão mostraram grande variabilidade de entendimento sobre o conceito de resultados. Vimos mais tarde que tivemos algumas inconsistências no material do experimento; havia algumas partes

em que não mencionamos este tipo de resultado e este pode ter sido o motivo de um resultado ruim.

**Tabela 20 - Expectativas e Resultados Experimento 1 – Coleta de Resultados**

Aspecto do Resultado	Expectativa	Experimento I
Texto Consistente	80%	53%
Nível de Suporte Consistente	90%	42%
Confiança Consistente	70%	53%

Apesar do baixo resultado médio de coleta de resultados, percebemos que em muitos casos os participantes coletaram todos os resultados explícitos dos artigos. Como podemos perceber na Figura 43, 50% das submissões coletaram 100 % dos resultados explícitos e 28% das submissões não coletaram nenhum resultado. Este resultado nos faz crer que outras variáveis podem estar influenciando o resultado, como: o tipo do estudo, a experiência, ou como o artigo foi escrito. Este experimento não foi projetado para analisar estas questões, portanto não podemos analisar os dados neste âmbito.

Artigo	Subject	Tipo de Estudo	Experiência como Participante em Experimento	Experiência Replicando Experimentos	Experiência Projetando Experimentos	Experiência no Tópico	Resultados Explícitos %	Resultados Implícitos %	%Resultados
1	S1	Experimento Controlado	0	0	0	0	100	0	85.71
2	S1	Estudo de Caso	0	0	0	1a	0	0	0.00
3	S2	Estudo de Caso	0	0	0	2	0	0	0.00
3	S4	Estudo de Caso	2 (2)	1B	0	2 (2)	0	0	0.00
4	S2	Experimento Controlado	0	0	0	2	0	0	0.00
5	S8	Estudo de Caso	2(1)	0	2(4)	2 (1)	50	0	50.00
5	S2	Estudo de Caso	0	0	0	0	0	0	0.00
5	S3	Estudo de Caso	1 a.b	0	0	1a/b	50	0	50.00
7	S4	Estudo de Caso	2 (2)	1B	0	2 (1)	12.5	33.33	18.18
8	S3	Estudo de Caso	1 a.b	0	0	0	20	0	20.00
8	S5	Estudo de Caso	0	0	0	0	100	0	100.00
8	S4	Estudo de Caso	2 (2)	1B	0	2 (2)	100	0	100.00
9	S5	Estudo de Caso	0	0	0	2	100	0	83.33
10	S5	Experimento Controlado	0	0	0	0	100	0	50.00
11	S6	Survey	0	1B	0	1b	0	40	33.33
12	S6	Survey	0	1B	0	1b	100	0	87.50
15	S8	Estudo de Caso	2(1)	0	2(4)	0	100	0	100.00
17	S1	Experimento Controlado	0	0	0	0	100	0	60.00
17	S6	Experimento Controlado	0	1B	0	0	100	0	60.00
18	S3	Experimento Controlado	1 a.b	0	0	2	100	0	47.27
18	S4	Experimento Controlado	2 (2)	1B	0	2 (3)	100	0	0.00

**Figura 43 - Resultados individuais para o experimento I**

### 5.2.6.2 Os participantes classificarão os artigos da mesma forma que o oráculo?

Para responder a esta pergunta, comparamos cada campo na descrição do contexto (objetivos, variáveis dependentes e independentes, categoria dos participantes, categoria das tarefas) preenchido pelo participante com a descrição que o oráculo fez para cada artigo. A Tabela 21 descreve nossas expectativas e os resultados que encontramos para cada campo..

Os resultados para o campo “objetivos” não foram os esperados. 37% das submissões estavam totalmente corretas, usando o GQM, 63% das respostas não estavam totalmente corretas, mas parcialmente. Algumas submissões descreveram o objetivo da tarefa ao invés do objetivo do estudo. Por exemplo: Em vez de “analisar técnicas da leitura”. O participante colocou “analisar requisitos”. Acreditamos que os principais fatores que influenciaram este resultado foram: o modelo do objetivo do GQM não foi seguido corretamente, e os participantes tenderam a escrever objetivos para a tarefa somente.

Sobre a coleta das variáveis, para variáveis dependentes 58% tiveram exatamente as mesmas variáveis encontradas pelo oráculo e 81% das submissões tiveram pelo menos as variáveis encontradas pelo oráculo. Se dividido pelo tipo de estudo temos o seguinte resultado: 72% das submissões tiveram exatamente as mesmas variáveis dependentes coletadas para experimentos controlados (5 de 7 submissões), 83% para estudos de caso (10 de 12 submissões) e 100% para surveys (2 de 2 submissões).

Para variáveis independentes, 32% das submissões tiveram exatamente as mesmas respostas encontradas pelo oráculo. 42% tiveram pelo menos as variáveis encontradas pelo oráculo. Se dividido pelo tipo de estudo temos o seguinte resultado: 72% das submissões tiveram exatamente as mesmas variáveis independentes coletadas para experimentos controlados (5 de 7 submissões), 25% para estudos de caso (3 de 12 submissões) e 50% para surveys (1 de 2 submissões).

Os resultados mostram que variáveis independentes são mais difíceis de extrair para os artigos que não são sobre experimentos controlados. E variáveis independentes são mais difíceis de reconhecer do que variáveis dependentes. O resultado não foi ainda o esperado,

---

mas os resultados para variáveis dependentes foram bem melhores do que para variáveis independentes. Um problema freqüente nas respostas para as variáveis dependentes era que os participantes tenderam a dizer que algumas variáveis eram variáveis dependentes quando estas eram independentes. E o erro mais comum na coleta das variáveis independentes foi que os participantes reconheceram algumas das variáveis mas não todas. Detectamos que a explanação que demos sobre as definições de variáveis dependentes e independentes não foi suficiente para o sucesso da extração.

Para o campo “participantes”, os resultados foram quase como o esperado. Aproximadamente 72% das submissões tiveram a mesma categoria encontrada pelo oráculo. Se dividido pelo tipo de estudo temos o seguinte resultado: 86% das submissões tiveram exatamente as mesmas categorias coletadas para experimentos controlados (6 de 7 submissões), 67% para estudos de caso (8 de 12 submissões) e 50% para surveys (1 de 2 submissões). Percebemos que houve certa dificuldade em definir os participantes para estudos quando estes não eram experimentos controlados.

A consistência entre os participantes em reconhecer as tarefas foi menor do que esperávamos. Mas chegou ao esperado em artigos nos quais o tipo de estudo era “experimento controlado”. Aproximadamente 62% das submissões tiveram a mesma categoria encontrada pelo oráculo. Se dividido pelo tipo de estudo temos o seguinte resultado: 86% das submissões tiveram exatamente as mesmas categorias coletadas para experimentos controlados (6 de 7 submissões), 50% para estudos de caso (6 de 12 submissões) e 50% para surveys (1 de 2 submissões).

**Tabela 21 - Expectativas e Resultados - Coleta de Informações de Contexto - Experimento 1**

<b>Variáveis de Contexto</b>	<b>Expectativas</b>	<b>Experimento I</b>
<b>Objetivos</b>	80%	37%
<b>Variáveis Dependentes</b>	80%	58%
<b>Variáveis Independentes</b>	70%	42%
<b>Participantes</b>	80%	72%
<b>Tarefas</b>	60%	62%

### 5.2.6.3 Qual é o tempo médio gasto em cada artigo?

Para responder a esta pergunta calculamos o tempo médio gasto em minutos para cada artigo, o número das páginas por artigo e se o artigo estava sendo lido pela primeira ou segunda vez.

Os participantes gastaram menos tempo do que esperado para os artigos que estavam sendo lidos pela primeira vez. Esperávamos que para os artigos lidos pela primeira vez os participantes gastariam entre 2.5 a 3.0 horas para ler e extrair a informação de um artigo, mas sete das oito submissões foram realizadas entre 1.5 a 3 horas e a média foi 8.1 minutos por página. Para quase todos os participantes, este foi o último artigo a ser lido. Este é um bom indicador de que a técnica não é muito intrusiva na leitura e também não requer muito mais tempo do que o normalmente utilizado para a leitura de artigos. Para os artigos que os participantes já tinham lido antes do experimento, esperávamos que os participantes gastariam entre 1 a 2 horas para ler e extrair a informação de um artigo e realmente 11 das 13 submissões foram realizadas entre 1 e 2 horas e a média foi de 7.22 minutos por página, mostrando que o tempo de extração dos artigos diminui depois que uma leitura do texto já foi realizada.

### 5.2.6.4 Questionário Pós-Estudo

Os dados qualitativos foram coletados usando um questionário de pós-estudo, que foi dividido em três seções principais:

- 1) Impressões sobre a técnica;
- 2) Impressões sobre os resultados;
- 3) Sugestões sobre como melhorar a técnica e o experimento.

Na primeira e segunda seção pedimos que os participantes indicassem sua opinião (discordo totalmente, discordo parcialmente, não tenho opinião, concordo parcialmente, concordo totalmente) para cada sentença abaixo e então justificar suas respostas:

Sentenças da primeira seção:

- 
- a) A técnica é útil para extrair a informação dos artigos.
- 85.71% (6 dos 7) dos participantes responderam que concordam parcialmente ou concordam totalmente com esta sentença. E os participantes acham que a técnica se torna mais fácil de usar após algumas aplicações em alguns artigos, indicando que há algum tipo de efeito de aprendizagem.
- b) Se eu tivesse mais treinamento eu poderia aplicar melhor a técnica.
- 71.43% (5 dos 7) dos participantes responderam que concordam de certa forma ou concordam totalmente com esta sentença. Alguns participantes sugeriram que talvez um exemplo concreto a priori ajudasse na confiança ao extrair a informação.

Sentenças da segunda seção:

- a) Usando a técnica eu pude extrair todos os resultados importantes do artigo.
- 100% dos participantes responderam que concordam, com algumas ressalvas, ou concordam totalmente com esta sentença.
- b) Os campos na informação do contexto capturam a maioria da informação que eu senti que eram importantes sobre o contexto experimental do estudo no artigo.
- 85.71% (6 dos 7) participantes responderam que concordam, com algumas ressalvas, ou concordam totalmente com esta sentença. Alguns participantes tiveram problemas em identificar os participantes do estudo porque a maioria dos artigos era estudos de caso.
- c) Se eu tivesse mais tempo, eu poderia aplicar melhor a técnica.
- 57.14% (4 dos 7) participantes responderam que concordam, com ressalvas, ou concordam totalmente com esta sentença.
- d) Se eu fosse generalizar resultados através dos múltiplos artigos, seria somente necessário olhar a informação extraída (isto é, eu não teria que ler outra vez os artigos).

- 57.14% (4 dos 7) participantes responderam que concordam, com ressalvas, ou concordam totalmente com esta sentença.

Nas sugestões sobre como melhorar a técnica e o experimento, os participantes expressaram sua satisfação sobre a experiência e a técnica, e a sugestão que mais apareceu foi sobre a necessidade de um treinamento mais profundo.

#### 5.2.6.5 Discussão Aberta

Após a análise dos resultados coletados, apresentamos os resultados aos participantes. Foi uma apresentação com uma discussão aberta sobre o experimento. Houve muitas introspecções interessantes que vieram dos participantes.

Discutindo sobre nossa questão de pesquisa, um dos participantes apontou que não havia nenhuma pergunta sobre a conformidade do processo já que não temos nenhuma informação sobre como o processo foi seguido. Por exemplo, não há nenhuma pergunta para se o participante executou diferentemente a técnica ao ler um artigo pela primeira vez ou ao ler um artigo que foi lido anteriormente para outros propósitos.

Sobre o projeto experimental, os participantes apontaram o fato de que eles eram todos inexperientes na técnica, o que acham que este fato pode ter tido efeito sobre os resultados. Um deles disse: “Eu sei que a primeira vez que eu apliquei a técnica eu realmente apliquei errado”. Já que só tínhamos dados destes participantes, isto parece ser uma falha no projeto. Durante esta discussão, também apontamos que era uma fraqueza do projeto que não consideramos tentar ter números aproximadamente iguais dos artigos nas várias categorias (estudos de caso, experimentos controlados e pesquisas de opinião). Também não balanceamos os tipos de artigos entre os participantes; por exemplo, devíamos ter tido uma regra na qual cada participante não pudesse fazer mais de um artigo descrevendo um estudo de caso.

Outro ponto que foi levantado nesta discussão foi a necessidade de saber a ordem de leitura dos artigos para compreender se existe um efeito de aprendizagem sobre os resultados.

---

Apesar de não o termos controlado, tínhamos as datas e o horário nos quais os participantes fizeram a extração, e então pudemos saber a ordem em que cada pessoa os leu.

Havia muita discordância sobre a definição de um resultado e como eles estavam descritos nos artigos. Alguns participantes disseram que alguns artigos tinham os resultados já de cara. Mas este problema era porque a nossa definição para resultado não estava bem explicitada no material do experimento. Tivemos uma discussão sobre as diferentes estratégias usadas:

- Coletar todos os resultados, mas eles tinham dúvidas sobre esta estratégia porque o texto da técnica os fazia pensar que eles deveriam coletar somente os resultados das hipóteses estabelecidas para o experimento.
- Em princípio coletar todos os resultados; depois deixar só o que era resultado importante do artigo e que trazia alguma contribuição substancial para a área (note que esta estratégia deixa de fora todas os resultados implícitos do estudo).
- Em princípio coletar só o que é importante e depois coletar todos os resultados auxiliares às hipóteses iniciais do experimento.

Falando sobre o modelo do contexto, a pergunta principal era: Como definir uma “tarefa” quando o artigo não relata um experimento controlado? Esta pergunta foi freqüente porque a maioria dos artigos eram estudos de caso e não havia nenhum executor ou tarefa específica a ser executada. Os estudos de caso geralmente consistiam de analisar os dados coletados previamente durante o desenvolvimento dos projetos.

Depois que fizemos esta breve discussão sobre o projeto e alguns conceitos, mostramos então os resultados, como descrito na Seção 5.2.6. Pedimos que os alunos comentassem sobre sua expectativa sobre cada questão que queríamos responder neste experimento; este comentário poderia conter causas e explicações sobre resultados finais:

- **O que é esperado para a consistência entre a extração dos resultados pelos participantes e pelo oráculo?**
  - Alguns dos participantes disseram 50%, outros pensaram que provavelmente mais do que isto. Eles fizeram uma boa previsão, foi basicamente o que aconteceu.

- Segundo os participantes, as porcentagens baixas para “resultados de outros artigos” foram provavelmente devido ao fato de que quando alguém lê pela segunda vez um artigo, gasta-se menos tempo preocupado com os trabalhos relacionados. As porcentagens baixas para resultados implícitos eram provavelmente devido à confusão sobre estilos diferentes para encontrá-los.
- Os participantes disseram ter problemas com a definição do suporte de confiança para os resultados, eles acharam que dependia muito de se eles haviam gostado do artigo ou não. Os participantes acharam que teriam tido 60% de consistência com o oráculo, e tiveram 53%, portanto, o resultado foi quase como esperado por eles.
- **O que é esperado para a consistência entre a extração do contexto pelos participantes e pelo oráculo?**
  - Os estudantes acreditam que deve ser consideravelmente elevado já que eles tiveram as aulas do professor Victor Basili sobre contextos experimentais também em Objetivos do GQM.
  - Para os objetivos GQM, se definirmos que parcialmente correto = errado, o resultado para este campo foi muito abaixo do esperado, os participantes admitiram que provavelmente confundiram objetivos da tarefa com os objetivos do estudo.
  - Os participantes tiveram uma confiança muito baixa sobre a extração das variáveis dependentes e independentes. Eles acharam que estava bem definido em classe, mas que sentiram dificuldades em identificá-las no texto. A expectativa de acerto para as variáveis independentes era maior do que para as variáveis dependentes. Para variáveis dependentes eles esperaram aproximadamente 30% a 50% da exatidão, mas foi mais elevada, chegando a 60%.
  - Sobre a classificação dos executores dos estudos (participantes), eles muitas vezes sentiram dificuldades em definir quem eram os participantes do estudo experimental no caso das pesquisas de opinião.
  - Sobre as tarefas, os participantes tiveram uma confiança muito baixa em suas respostas para este campo. Esperaram aproximadamente 70% a 80% da exatidão. O

---

resultado foi ainda menor, sendo 60% de exatidão. Para experimentos controlados o resultado chega aos 80% o que é consistente com a expectativa dos participantes.

### 5.2.7 Ameaças à Validade

Uma pergunta fundamental a respeito dos resultados de um estudo experimental é saber quão válidos são os resultados. Nesta seção discutimos algumas ameaças às validades internas, de construção e externas.

A validade interna tem a ver com a defesa contra fontes de equívocos no projeto de pesquisa, que afetaria o processo causa-efeito que está sendo estudado, introduzindo variáveis imprevistas. Vemos o efeito *Hawthorne* (expectativa do experimentador) como um possível problema de validade interna no estudo. Acreditamos que há uma possibilidade de que as expectativas ou as ações do pesquisador possam contaminar os resultados, já que o projetista era o oráculo e também o avaliador dos resultados.

A maturação é outra possível ameaça à validade interna; o treinamento que definimos para o estudo não foi completo o suficiente para cobrir todos os aspectos da técnica. Além disto, não controlamos a ordem em que os participantes analisaram os artigos; assim não pudemos analisar os efeitos de aprendizagem na efetividade dos resultados.

A validade da construção, chamada às vezes também validade fatorial, tem a ver com a lógica dos itens que compreendem medidas de conceitos sociais. O projeto experimental do Experimento I foi mono-método. Usamos um único método de coleta, o que poderia trazer equívocos, mas usamos a aleatorização dos artigos e da seleção dos participantes.

A validade externa tem a ver com a possibilidade de equívocos no processo de generalização das conclusões de uma amostra para uma população, a outras populações, a outros ajustes, e/ou a outros períodos de tempo. Dos sete participantes, somente um teve a experiência precedente com projeto de estudos experimentais; os resultados talvez fossem diferentes para outro grupo de participantes. Um dado importante a fortalecer a validade externa do estudo é que os artigos que usamos eram de periódicos diferentes baseados em tópicos diferentes da Engenharia de Software.

### **5.3 Avaliação experimental em artigos de teste.**

Pode ser impossível realizar um experimento sem determinadas ameaças. Além disto, algumas questões ainda não estavam claras depois do primeiro experimento. E gostaríamos de revisar alguns problemas no experimento e ver qual seria o resultado final. A lista de artigos utilizada era o principal problema pois, por serem de vários tópicos, trazia uma dificuldade maior em encontrar participantes com experiência suficiente para entender os artigos. Decidimos então realizar um experimento somente com artigos de teste. E também somente com artigos que tratavam experimentos controlados.

Também, após a realização do experimento descrito na seção anterior, fizemos algumas alterações no material e também na técnica para resolver alguns problemas que encontramos ao realizarmos o experimento. O objetivo para este experimento foi o mesmo do anterior, os experimentadores e a coleta de dados também.

Além das perguntas para o ultimo experimento, nessa repetição tivemos a seguinte pergunta específica:

- Os resultados para este experimento serão melhores do que os resultados do experimento anterior?

#### **5.3.1 Experimentadores**

Pesquisadores na Universidade de Maryland que tinham criado a técnica e eram conseqüentemente peritos em seu uso. O projetista da técnica é o oráculo para as respostas.

#### **5.3.2 Participantes.**

Os participantes eram estudantes de pós-graduação na Universidade Salvador, matriculados em uma turma de Engenharia de Software Experimental em 2006. Neste estudo, havia 22 participantes. Como será descrito na seção do procedimento, cada participante extraiu informações de 2 artigos. Os participantes 10, 19, 20, 21 e 22 desistiram do experimento.

Para este experimento coletamos a mesma informação que no primeiro experimento, e além da experiência coletamos o nível de conhecimento em inglês, já que desta vez o experimento estava sendo executado no Brasil, mas todo o material estava em inglês. A Tabela 22 mostra os perfis para cada participante. A escala usada para o nível da experiência era a seguinte:

**0 = Nenhuma experiência.**

**1 = Experiência somente em sala de aula.**

**Deve-se indicar se:**

**a) Aprendeu somente os conceitos em classe;**

**b) Usou o conceito em um trabalho de classe;**

**c) Usou o conceito em um projeto.**

**2 = Experiência Profissional.**

**Indique o número de projetos nos quais você executou a atividade.**

**Tabela 22- Experiência dos participantes do Experimento II**

Participante	Número de anos de Experiência em Engenharia de Software	Experiência em Participar de um Experimento	Experiência em Replicar Experimentos	Experiência em Projetar Experimentos	Nível de Inglês
S1	6	0	0	0	Baixo
S2	9	0	1.c	1.c	Alto
S3	5	2 (1)	1	1	Alto
S4	11	0	0	0	Médio
S5	7	2	0	0	Alto
S6	6	0	0	0	Médio
S7	4	0	0	0	Baixo
S8	4	0	0	1	Médio
S9	2	0	1.a	1.c	Médio
S11	5	1.a	0	0	Alto
S12	10	0	0	0	Médio
S13	15	0	0	0	Médio
S14	7	0	0	0	Baixo
S15	4	0	1.c	2(8)	Médio
S16	4	0	0	0	Médio
S17	20	0	0	0	Médio
S18	10	2	2	2	Médio

### 5.3.3 Materiais

A lista de material usado no experimento foi basicamente o mesmo usado no experimento anterior, com algumas melhorias no treinamento, na técnica e nos formulários de coleta, e a lista de artigos foi diferente. Foram usados nove artigos (contando o artigo do treinamento), oito deles eram experimentos controlados. Os artigos são um subconjunto dos artigos analisados por Juristo e outros [100]:

1. Victor R. Basili, Richard Selby, Comparing the Effectiveness of Software Testing Strategies, *IEEE Transactions on Software Engineering*, 13(12):1278-1296, December 1987.
2. Elaine J. Weyuker. The cost of data flow testing: An empirical study. *IEEE Transactions on Software Engineering*, 16(2): 121–128, February, 1990.
3. James M. Bieman, Janet L. Schultz. An empirical evaluation (and specification) of the all-du-paths testing criterion. *IEE Software Engineering Journal*,7(1):43–51, January,1992.
4. Phyllis G. Frankl, Stewart N. Weiss. An experimental comparison of the Effectiveness of branch testing and data flow testing. *IEEE Transactions on Software Engineering*, 19(8): 774–787, August, 1993.
5. Monica Hutchins , Herb Foster , Tarak Goradia , Thomas Ostrand. Experiments on the Effectiveness of data-flow and control-flow-based test adequacy criteria. *Proceedings of the 16th International Conference on Software Engineering*. 191–200, Sorrento, Italy, May, 1994.
6. Phyllis G. Frankl, Oleg Iakounenko. Further empirical studies of test Effectiveness. In *Proceedings of the ACM SIGSOFT International Symposium on Foundations on Software Engineering*, 153–162, Lake Buena Vista, Florida, USA, November, 1998.
7. Glenford J. Myers. A controlled experiment in program testing and code walkthroughs/inspections. *Communications of the ACM*, 21(9): 760–768, September, 1978.
8. Erik Kamsties, Christopher M. Lott. An empirical evaluation of three defect-detection techniques. *Proceedings of the Fifth European Software Engineering Conference*, pages 362-383, Sitges, Spain, 1995.

Treinamento:

9. Murray Wood, Marc Roper, Andrew Brooks, James Miller. Comparing and combining software defect detection techniques: A replicated empirical study. *ACM SIGSOFT Software Engineering Notes*, 22 (6): 262 – 277, November, 1997.

### 5.3.4 Procedimento

Um questionário sobre experiência foi distribuído aos participantes para coletar a informação relevante sobre suas experiências no desenvolvimento de software. Uma cópia deste questionário pode ser encontrada no Apêndice A. O experimento foi uma das atividades da disciplina. Ministramos o treinamento aos estudantes por uma hora e trinta minutos (01h30 min) e os participantes aplicaram a técnica em um artigo específico e muito fácil de encontrar toda a informação que a técnica busca. A cada um dos participantes foram atribuídos dois outros artigos. Cada artigo foi lido por quatro participantes. O projeto experimental deste experimento é descrito na Tabela 23.

**Tabela 23 - Projeto Experimental do Experimento II**

Primeiro Dia																	
Parte A 00h30min	Formulário de consentimento e formulário de experiência dos participantes																
Parte B 01h30min	Treinamento: Visão Geral da Técnica e aplicação da técnica em um artigo																
Em casa																	
Parte C 4 horas	Particip.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Artigos	1,5	2,6	3,7	4,8	1,8	5,2	3,6	7,4	7,1	8,2	5,3	6,4	6,1	2,7	8,3	4,5
Parte D 30 min	Formulário de Pós-estudo.																

Os participantes tiveram uma semana para analisar os dois artigos em casa. Após ter realizado a análise nos artigos, cada participante respondeu a um questionário pós-estudo com perguntas sobre suas impressões sobre a técnica e os resultados e sugestões sobre como podemos melhorar a SecESE. Cinco estudantes desistiram do experimento.

Finalmente, depois que os participantes terminaram a análise e todos os questionários, uma discussão em classe foi realizada, como no primeiro experimento. No final obtivemos dados de 17 participantes, com 34 submissões.

### 5.3.5 Resultados/Lições Aprendidas

Os resultados são apresentados nessa seção, e seguidos de uma discussão. Os resultados apresentados são baseados em três questões principais:

- 1) Os participantes encontrarão os mesmos resultados nos artigos que o oráculo?
- 2) Os participantes classificarão os artigos da mesma forma que o oráculo?
- 3) Qual é o tempo médio gasto em cada artigo?

#### 5.3.5.1 Os participantes encontrarão os mesmos resultados que o oráculo?

Para responder a esta pergunta, comparamos cada resultado que o oráculo encontrou em cada artigo com os resultados encontrados por cada participante.

Apesar de termos modificado a técnica nos pontos fracos detectados no experimento anterior, os resultados ainda não foram mais eficientes. De fato, os resultados foram muito piores do que os resultados do primeiro experimento (53% dos resultados). Os participantes encontraram em média: **27% dos resultados testados e 17% dos “resultados de outros artigos”**. A maioria dos resultados encontrados eram explícitos. A maioria dos resultados implícitos não foi encontrada pelos participantes. Os participantes encontraram: **45% dos resultados explícitos e 9% dos resultados implícitos**.

Percebemos nas respostas que havia um problema grande com o inglês; muitas respostas continham sentenças incompletas que não faziam sentido, e então eram invalidadas para o estudo. Somente quatro dos dezessete participantes disseram ter um nível inglês elevado. E sobre a experiência em estudos experimentais, quase todos não tiveram nenhuma experiência, em nenhum nível. Sobre a experiência no tópico tratado no artigo, somente um participante tinha experiência. No primeiro experimento supusemos que a experiência talvez não fosse um fator que afetasse os resultados, mas aparentemente afeta os resultados.

A avaliação do participante sobre o que é um inglês bom talvez não esteja em uma mesma base da avaliação para todos eles. Vale a pena observar que as melhores respostas que tivemos foi de um estudante com nível de inglês médio; ele coletou em média 80% dos resultados relevantes (explícitos e implícitos) e 100% dos resultados explícitos. Uma

característica importante é que ele é o único participante que tem a experiência em projetar estudos experimentais profissionalmente; ele relatou que tinha projetado 8 estudos experimentais na indústria até então.

Como no experimento I, apesar do baixo resultado médio de coleta de resultados, percebemos que em muitos casos os participantes coletaram todos os resultados explícitos dos artigos. Como podemos perceber na Figura 44, 30% das submissões coletaram mais de 70% dos resultados explícitos e 30% das submissões não coletaram nenhum resultado. Este resultado nos faz crer que outras variáveis podem estar influenciando o resultado (a experiência em experimentação, a experiência no tópico ou como o artigo foi escrito - dificuldade do artigo).

Artigo	Subject	Experiência como Participante em Experimento	Experiência Replicando Experimentos	Experiência Projetando Experimentos	Experiência no Tópico	Nível de Inglês	Dificuldade do Artigo	Resultados Explícitos %	Resultados Implícitos %	%Resultados
1	S17	0	0	0	0	Médio	n/a	25.93	0.00	11.67
1	S9	0	1.a	1.c	0	Médio	Médio	100.00	0.00	50.00
1	S5	2	0	0	0	Alto	Baixo	36.00	0.00	15.00
1	S13	0	0	0	0	Médio	Médio	0.00	0.00	0.00
2	S18	2	2	2	0	Médio	Médio	33.33	0.00	11.11
2	S2	0	1.c	1.c	1.a	Alto	Baixo	66.67	0.00	22.22
2	S6	0	0	0	1.B	Médio	Médio	100.00	100.00	100.00
2	S14	0	0	0	0	Baixo	Alto	28.57	0.00	14.29
3	S3	2 (1)	1	1	1	Alto	Alto	0.00	0.00	0.00
3	S7	0	0	0	0	Baixo	Médio	100.00	0.00	50.00
3	S11	1.a	0	0	1.B	Alto	Médio	0.00	0.00	0.00
3	S15	0	1.c	2(8)	0	Médio	n/a	100.00	50.00	88.24
4	S4	0	0	0	1.a	Médio	Alto	37.50	0.00	20.00
4	S8	0	0	1	1.a	Médio	Baixo	71.43	42.86	57.14
4	S12	0	0	0	1.B	Médio	Médio	0.00	0.00	0.00
4	S16	0	0	0	0	Médio	Médio	0.00	0.00	0.00
5	S1	0	0	0	0	Baixo	n/a	0.00	0.00	0.00
5	S11	1.a	0	0	1.a	Alto	Médio	0.00	0.00	0.00
5	S6	0	0	0	1.B	Médio	Baixo	33.33	0.00	11.11
5	S16	0	0	0	0	Médio	Médio	75.00	0.00	50.00
6	S17	0	0	0	0	Médio	n/a	0.00	0.00	0.00
6	S2	0	1.c	1.c	1.a	Alto	Médio	50.00	0.00	50.00
6	S7	0	0	0	0	Baixo	Alto	0.00	0.00	0.00
6	S13	0	0	0	0	Médio	Médio	28.00	0.00	11.67
6	S12	0	0	0	1.B	Médio	Médio	25.00	0.00	13.33
7	S18	2	2	2	0	Médio	Alto	85.71	0.00	42.86
7	S3	2 (1)	1	1	1	Alto	Baixo	42.86	71.43	57.14
7	S9	0	1.a	1.c	0	Médio	Médio	100.00	0.00	41.67
7	S8	0	0	1	2 (3)	Médio	Baixo	100.00	14.29	60.00
7	S14	0	0	0	0	Baixo	Alto	66.67	0.00	22.22
8	S4	0	0	0	0	Médio	Alto	0.00	0.00	0.00
8	S5	2	0	0	0	Alto	Médio	53.85	0.00	41.18
8	S15	0	1.c	2(8)	0	Médio	n/a	100.00	50.00	75.00

**Figura 44 - Resultados Individuais Experimento II**

### 5.3.5.2 Os participantes classificarão os artigos da mesma forma que o oráculo?

Para responder a esta pergunta, comparamos cada campo na descrição do contexto preenchido pelo participante com a descrição do contexto do oráculo. O processo foi o mesmo do primeiro experimento.

A Tabela 24 descreve nossas expectativas e os resultados que obtivemos para os campos: objetivos, variáveis dependentes e independentes, categoria de participantes e categoria das tarefas.

Os resultados para o campo “objetivos” foram melhores do que esperado; 42% estavam totalmente corretos, usando o GQM e 88% das respostas não estavam totalmente corretas, mas parcialmente. Os problemas principais foram que o modelo do objetivo GQM não foi seguido corretamente. O fato que os participantes tenderam a escrever objetivos para a tarefa somente, não ocorreu desta vez. A explanação para os resultados melhores pode ser o treinamento no qual focamos um pouco mais neste modelo e como preenchê-lo corretamente.

Sobre a coleta das variáveis, os resultados reafirmam o resultado de que “para experimentos controlados as variáveis dependentes e independentes são igualmente difíceis de reconhecer”. Neste experimento, os participantes encontraram 62% das variáveis dependentes e 60% das variáveis independentes, um pouco menos do que os resultados precedentes para experimentos controlados (71% para variáveis dependentes e também para as variáveis independentes). O problema freqüente do último experimento, que os participantes tenderam a dizer que algumas variáveis eram variáveis dependentes e não eram, não ocorreu muito desta vez. O erro mais comum na coleta das variáveis independentes foi que os participantes reconheceram somente algumas variáveis.

Para o campo “participantes”, os resultados foram quase como esperávamos. 76% das submissões tiveram a mesma categoria encontrada pelo oráculo. Considerando que todos os estudos eram experimentos, esperávamos resultados melhores, mas considerando que todos os artigos são de teste, e nem sempre existem “participantes”; além disto, algumas vezes os autores chamavam os “produtos de trabalho” de “subjects”, mesma palavra em inglês para se

---

referir a “participantes”, o que confundiu alguns participantes; isto é ainda um agravante maior neste experimento porque os participantes não são nativos da língua inglesa.

Para o campo “tarefas”, o fato de que os artigos eram artigos de teste foi um fator para estabelecermos que só esperávamos que 60% das respostas fossem consistentes com o oráculo. O resultado foi bem perto das expectativas, 53% das submissões tiveram a mesma categoria encontrada pelo oráculo. O que acontece nestes artigos é que em quase todos eles não existem participantes e nem tarefas a serem executadas por humanos, mas por alguns programas; nem sempre é claro nestes artigos que tarefas são estas.

Os campos “produtos de trabalho” e “instrumentação” não estavam no modelo quando executamos o último experimento, essas inclusões foram avaliadas pela primeira vez e os resultados estiveram muito próximos do que se esperava deles. Para produtos de trabalho, 73% das submissões tiveram os mesmos produtos de trabalho encontrados pelo oráculo, e 74.24% das submissões tiveram a mesma instrumentação encontrada pelo oráculo.

**Tabela 24 - Expectativas e Resultados - Coleta de Informações de Contexto no Experimento II**

<b>Variáveis de Contexto</b>	<b>Expectativas</b>	<b>Experimento I</b>	<b>Experimento II</b>
<b>Objetivos</b>	80%	37%	42%
<b>Variáveis Dependentes</b>	80%	58%	62%
<b>Variáveis Independentes</b>	70%	42%	60%
<b>Participantes</b>	80%	72%	76%
<b>Tarefas</b>	60%	62%	53%
<b>Produtos de Trabalho</b>	80%	Não Existia	73%
<b>Instrumentação</b>	80%	Não Existia	74%

#### 5.3.5.3 Qual é o tempo médio gasto em cada artigo?

Para responder a esta pergunta calculamos o tempo médio gasto para cada artigo, coletamos o número das páginas por artigo e se era o primeiro ou o segundo artigo que eles estavam revisando. Este resultado é incomparável com o resultado do experimento anterior, já que é claro que nativos em inglês lerão os artigos muito mais rapidamente.

Esperávamos que o tempo gasto na extração a informação de documentos diminuiria do primeiro para o segundo artigo. Esperávamos que os participantes gastariam entre 2.5 a 3.0 horas para extrair as informações de um artigo, O tempo médio gasto por artigo foi de 2.5 horas com uma média de 13.89 minutos por página. Para o segundo artigo a ser analisado, esperávamos que os participantes gastariam entre 1.5 a 2.5 horas para extrair as informações de um artigo, O tempo médio gasto por artigo foi de 2.6 horas com uma média de 14.48 minutos por página. Os resultados mostram portanto que não existe uma diminuição no tempo gasto entre a primeira extração e a segunda; acreditamos porém, que a questão da língua pode ser um fator importante no efeito dos resultados.

#### 5.3.5.4 Questionário de Pós-Estudo

Assim como no experimento anterior, os dados qualitativos foram coletados usando um questionário de pós-estudo, uma apresentação feita após a análise dos dados coletados, com uma discussão aberta.

O questionário de pós-estudo foi dividido em três seções principais:

- 1) Impressões sobre a técnica;
- 2) Impressões sobre os resultados;
- 3) Sugestões sobre como melhorar a técnica e o experimento.

Na primeira e segunda seção solicitamos aos participantes que indicassem seu acordo com cada sentença abaixo e então justificar suas respostas:

Sentenças da primeira seção:

- a) A técnica é útil para extrair a informação dos artigos.
  - 100% dos participantes responderam que concordam de certa forma ou concordam totalmente com esta sentença.
- b) Se eu tivesse mais treinamento eu poderia aplicar melhor a técnica.

- 
- 82.23% dos participantes responderam que concordam de certa forma ou concordam totalmente com esta sentença. Alguns sugeriram que fizéssemos o treinamento com mais de um artigo, e que tivéssemos mais tempo de treinamento.

Sentenças da segunda seção:

- a) Usando a técnica eu pude extrair todas os resultados importantes do artigo.
  - 100% dos participantes responderam que concordam, com algumas ressalvas, ou concordam totalmente com esta sentença.
- b) Os campos na informação do contexto capturam a maioria da informação que eu senti que eram importantes sobre o contexto experimental do estudo no artigo.
  - 88.23% dos participantes responderam que concordam, com algumas ressalvas, ou concordam totalmente com esta sentença.
- c) Se eu tivesse mais tempo, eu poderia aplicar melhor a técnica.
  - 76.47% dos participantes responderam que concordam, com ressalvas, ou concordam totalmente com esta indicação. Um fato importante é que a maior parte deles trabalha na indústria; portanto, eles disseram ter pouco tempo em uma semana para realizar a tarefa a contento.
- d) Se eu fosse generalizar resultados através dos múltiplos artigos, seria somente necessário olhar a informação extraída (isto é, eu não teria que ler outra vez os artigos).
  - 70.59% dos participantes responderam que concordam com ressalvas ou concordam totalmente com esta indicação.

Nas sugestões sobre como melhorar a técnica e o experimento, os participantes expressaram sua satisfação sobre a experiência e a técnica, e a sugestão que mais apareceu foi sobre a necessidade de um treinamento maior e com exemplos de aplicação da técnica nos artigos. Alguns participantes também comentaram sobre a dificuldade com a língua inglesa.

## 5.4 Entrevista com Especialistas

O principal objetivo da entrevista foi entender o processo usado no artigo “Reviewing 25 years of testing technique experiments” [100]. Entrevistamos duas das autoras do artigo: Natalia Juristo e Sira Vegas (Especialista I e II). Elas foram contatadas durante o ISESE 2006 (International Symposium on Empirical Software Engineering). As entrevistas tiveram duração de uma hora mais ou menos. Na sala, além das entrevistadas estavam os seguintes pesquisadores: Daniela Cruzes, Manoel Mendonça, Forrest Shull, Victor Basili. A transcrição completa da entrevista está no Apêndice B.

A entrevista foi semi-estruturada e as questões previstas para a entrevista foram as seguintes:

1. Quais foram os critérios de seleção dos artigos? Especifique critérios para a inclusão e a exclusão dos artigos.
2. Como foi a divisão dos artigos? Cada uma de vocês revisou o artigo individualmente ou em pares? Existia uma ordem pré-definida para a leitura dos artigos?
3. Havia um processo específico para ler cada artigo?
4. Havia uma informação específica que vocês procuravam em cada artigo?
5. Você tem a idéia do tempo gasto em cada artigo?
6. O que torna um artigo difícil ou fácil de ler?
7. Quanto sua experiência com teste ajudou na análise?
8. Quanto sua experiência com experimentação ajudou na análise?
9. Para você, qual é a definição de “resultado”? Um resultado válido tem sempre que ser baseado em dados?
10. Quando você estava analisando um artigo, você tentou extrair sempre todos os resultados que ele poderia ter, ou você teve um objetivo específico nos tipos de resultados que você quis coletar?
11. Como você tratou resultados de outros artigos? Você os considerou em sua análise?
12. Você leu o artigo inteiro ou apenas algumas seções específicas para procurar pela informação que você necessitava?

13. Para avaliar as limitações. Vocês usaram uma abordagem específica para escolher o que seria analisado? Por exemplo: tipo da análise, da categoria de assuntos etc. Se sim, você usou esta abordagem para todos os artigos?

14. Sabemos que vocês extraíram as variáveis dependentes dos estudos e as técnicas de testes usadas em cada estudo. Você acredita ser importante avaliar as variáveis independentes dos estudos?

15. Sobre os resultados práticos, vocês tiveram um processo para validá-los? Como contagem de votos etc.?

As entrevistas produziram resultados qualitativos interessantes, além de um processo de reflexão interessante. Pudemos em alguns momentos avaliar a SecESE, verificando que viável e útil; principalmente, pudemos perceber que a SecESE resolve muitos dos problemas que as autoras enfrentaram nas revisões que já fizeram. Extraímos algumas frases das entrevistas que chamaram mais a nossa atenção. Algumas destas frases são também hipóteses que já havíamos detectado, outras são confirmações para algumas de nossas opiniões.

<b>Extraído da Entrevista</b>	<b>Conclusões</b>
Para a Especialista I, é importante revisar as referências bibliográficas em busca de novas fontes de artigos. Para a Especialista II, não vale a pena despende esforço nesta tarefa porque o número de artigos incluídos nesta tarefa é muito pequeno.	Revisar ou não as referências em busca de novas referências bibliográficas depende do tópico em questão, mas nas experiências que tivemos até hoje, nos faz ter a mesma hipótese que a Especialista II
A revisão em pares é importante para a análise. Elas acreditam que esta é uma boa prática e acredita ser importante ter mais de uma pessoa lendo o artigo.	Precisamos realizar um experimento para avaliar esta questão. A nossa hipótese é de que uma vez que se tem uma técnica de extração como a que definimos, a necessidade de se ter mais de uma pessoa na extração diminui. Principalmente quando o analista tem experiência com experimentação e com o domínio dos artigos em revisão.
A dificuldade de extração nos artigos depende muito do artigo.	Claramente existem artigos mais fáceis e mais difíceis. Mas ainda não tivemos como comprovar a relação causa-efeito da dificuldade do artigo em relação à efetividade da extração das informações.

Existem artigos bons e ruins. Artigos completos e incompletos. Os piores casos são os artigos incompletos, porque você permanece lendo o artigo para ver se a informação está lá receosa de que ela está lá, mas você não consegue encontrá-la.	Tivemos esta situação em alguns artigos. E concordamos com a especialista.
Informações de contexto são mais fáceis de extrair do artigo do que os resultados. É mais uma questão de tempo para extrair algumas informações de contexto do que de conhecimento requerido.	Esta hipótese foi confirmada em nossos experimentos e nos motiva a investir na automação da extração de informações de contexto.
O tempo de extração das informações dos artigos depende muito do artigo.	De fato, usando a técnica de extração proposta neste trabalho, percebemos que o tempo de extração dos artigos é quase uniforme, mas acreditamos que esta uniformidade foi imposta pelos participantes dos experimentos por estabelecer implicitamente que não gastariam muito tempo em cada artigo.
Artigos que não explicitam os resultados claramente precisam ser interpretados. Isto gera margens a interpretações subjetivas o que altera e compromete a qualidade da análise assim como a sua repetibilidade.	Apesar de não termos executado experimentos que avaliem esta hipótese, concordamos com as especialistas.
As especialistas acreditam que para agregar estudos, a pessoa necessita conhecer o tópico. Talvez, tão importante ou mais que a experiência com experimentação.	Apesar de não possuímos dados experimentais que apoiem a hipótese, pudemos perceber que este fenômeno parece ser real quando realizamos os estudos de viabilidade.
Para a extração das informações, o melhor caso é quando o leitor é um perito em experimentação e no assunto que está revisando. Para Especialista I a experiência em experimentação foi ainda mais importante do que a experiência em teste. Para a Especialista II, ao ler e ao extrair a informação do artigo, o conhecimento do domínio não é tão importante quanto a experiência com experimentação.	Esta é uma questão importante. Decidimos projetar um experimento para analisar estas questões de experiência (vide Seção 5.5). Os resultados não indicam a priori diferença estatística na efetividade dos participantes nos dois grupos: experientes em experimentação e experientes no domínio.
Para Especialista II é necessário ter-se um modelo para buscar por informações de contexto	Concordamos, e por isto definimos os modelos para extração de informações e contexto e também para os resultados.

---

## 5.5 Experimento com Especialistas

Após a execução dos dois experimentos e a entrevista com os especialistas, vimos que precisávamos analisar com mais profundidade a influência da experiência na leitura e extração das informações dos artigos. Assim, neste experimento tivemos as seguintes perguntas específicas:

1. Os participantes encontrarão os mesmos resultados nos artigos que o oráculo?
2. Os participantes classificarão os artigos da mesma forma que o oráculo?
3. Qual o efeito da experiência dos participantes nos resultados?
4. Os resultados para este experimento serão melhores do que os resultados dos experimentos anteriores?

### 5.5.1 Experimentadores

Pesquisadores na Universidade de Maryland que tinham criado a técnica e eram consequentemente peritos em seu uso. O projetista da técnica é o oráculo para as respostas.

### 5.5.2 Participantes.

Os participantes eram pesquisadores experientes na área de experimentação ou na área de teste. Dividimos os participantes em dois grupos: os experientes em experimentação e os experientes em teste. Os participantes experientes em experimentação possuem algum conhecimento de teste, mas não são especialistas na área. Já os experientes em teste não são tão experientes em experimentação, mas têm um certo conhecimento. Para este experimento coletamos a mesma informação que no primeiro experimento. A Tabela 25 mostra os perfis para cada participante. A escala usada para o nível da experiência era a seguinte:

**0 = Nenhuma experiência.**

**1 = Experiência somente em sala de aula. Deve-se indicar se:**

- a) Aprendeu somente os conceitos em classe;
- b) Usou o conceito em um trabalho de classe;
- c) Usou o conceito em um projeto.

**2 = Experiência Profissional. Indique o número de projetos nos quais você executou a atividade.**

**Tabela 25 - Experiência dos Participantes - Experimento III**

Participante	Número de anos de Experiência em ES	Experiência em Participar de um Experimento	Experiência em Replicar Experimentos	Experiência em Projetar Experimentos	Especialista em:
S1	37	0	2 (muitos)	2 (muitos)	Experimentação
S2	15	1C	2 (2)	2(4)	Experimentação
S3	16	2(2)	2(2)	2(4)	Experimentação
S4	15	0	0	0	Teste
S5	30	0	0	0	Teste
S6	19	1C	0	2(1)	Teste

### 5.5.3 Materiais

O material usado no experimento foi basicamente o mesmo usado no experimento anterior, com algumas melhorias no treinamento, na técnica e nos formulários de coleta, e também a lista de artigos foi diferente. Foram usados dois artigos (um para o treinamento e um para a extração). Ambos os estudos eram experimentos controlados. Os artigos são um subconjunto dos artigos analisados por Juristo e outros [100]:

Análise:

1. Erik Kamsties, Christopher M. Lott. An empirical evaluation of three defect-detection techniques. *Proceedings of the Fifth European Software Engineering Conference*, pages 362-383, Sitges, Spain, 1995.

Treinamento:

2. Murray Wood, Marc Roper, Andrew Brooks, James Miller. Comparing and combining software defect detection techniques: A replicated empirical study. *ACM SIGSOFT Software Engineering Notes*, 22 (6): 262 – 277, November, 1997.

### 5.5.4 Procedimento

Distribuímos um questionário sobre experiência aos participantes para coletar informações sobre suas experiências. Depois realizamos um treinamento e fizemos uma discussão individual com cada participante. Após esta discussão eles aplicaram a técnica no artigo.

---

Cada participante então respondeu a um questionário do estudo pós-estudo com perguntas sobre suas impressões sobre a técnica e os resultados e, sugestões sobre como podemos melhorar a SecESE. O projeto experimental deste experimento é descrito na Tabela 23.

**Tabela 26 - Projeto Experimental do Experimento III.**

Projeto Experimental	
Parte A	Formulário de consentimento e formulário de experiência dos participantes
Parte B	Treinamento: Visão Geral da Técnica e aplicação da técnica em um artigo. Discussão individual sobre a extração das informações no artigo de treinamento.
Parte C	Aplicação da técnica em um Artigo.
Parte D	Preenchimento do formulário de Pós-estudo.

### 5.5.5 Resultados/Lições Aprendidas

Os resultados são apresentados nessa seção e então uma discussão segue. Os resultados apresentados são baseados em três questões principais:

- 1) Os participantes encontrarão os mesmos resultados nos artigos que o oráculo?
- 2) Os participantes classificarão os artigos da mesma forma que o oráculo?
- 3) Qual o efeito da experiência dos participantes nos resultados?
- 4) Os resultados para este experimento serão melhores do que os resultados dos experimentos anteriores?

#### 5.5.5.1 Os participantes encontrarão os mesmos resultados que o oráculo?

Para responder a esta pergunta, comparamos cada resultado que o oráculo encontrou em cada artigo com os resultados encontrados por cada participante. A planilha usada para coletar os dados foi a mesma usada no último experimento. A Tabela 28 descreve nossas expectativas e os resultados finais. 66% dos resultados foram coletados; os resultados deste experimento foram melhores do que os resultados do experimento I e II. É importante notar que este artigo foi o mais difícil já analisado até hoje para a base de evidências criada durante

este trabalho. Mas também, os participantes foram os mais qualificados que já participaram das avaliações.

**Tabela 27 - Resultados do Experimento III - Dividido por Experiência**

Participante	Especialista em:	% de Resultados Encontrados	Média por Especialidade
S1	Exp.	65%	73.33%
S2	Exp.	90%	
S3	Exp.	74%	
S4	Teste	58%	58.33
S5	Teste	79%	
S6	Teste	38%	

**Tabela 28 - Expectativas e Resultados Experimento III**

Aspecto do Resultado	Expectativa	Experimento I	Experimento II	Experimento III
Texto Consistente	80%	53%	27%	66%

#### 5.5.5.2 Os participantes classificarão os artigos da mesma forma que o oráculo?

Para responder a esta pergunta, comparamos cada campo na descrição do contexto preenchido pelo participante com a descrição do contexto do oráculo. O processo foi o mesmo que do primeiro experimento.

A Tabela 29 descreve nossas expectativas e os resultados que obtivemos para os campos: objetivos, variáveis dependentes e independentes, categoria de participantes, categoria das tarefas, produtos de trabalho e instrumentação. Quase todos os campos foram coletados como esperados, porém, surpreendeu-nos a questão de que as variáveis dependentes e independentes não tenham sido completamente coletadas.

E também no campo “tarefas”, somente 2 dos 6 participantes preencheram este campo como esperado. Todos os outros incluíram vários detalhes do projeto experimental que estava correto sobre a tarefa, mas que não contemplavam o modelo que definimos para a

---

formalização das tarefas, neste caso acreditamos que a experiência dos participantes teve um efeito negativo na coleta destas informações específicas.

**Tabela 29 - Expectativas e Resultados - Informações de Contexto - Experimento III**

<b>Variáveis de Contexto</b>	<b>Expectativas</b>	<b>Experimento I</b>	<b>Experimento II</b>	<b>Experimento III</b>
<b>Objetivos</b>	80%	37%	42%	100%
<b>Variáveis Dependentes</b>	80%	58%	62%	63%
<b>Variáveis Independentes</b>	70%	42%	60%	63%
<b>Participantes</b>	80%	72%	76%	100%
<b>Tarefas</b>	60%	62%	53%	33%
<b>Produtos de Trabalho</b>	80%	Não Existia	73%	100%
<b>Instrumentação</b>	80%	Não Existia	74%	100%

### 5.5.5.3 Qual o efeito da experiência dos participantes nos resultados?

A Tabela 27 mostra que os resultados dos especialistas em experimentos foram um pouco melhores que os resultados dos especialistas em teste, mas o teste estatístico para as diferenças entre as médias diz que a diferença entre as medidas das duas amostras não é significativa. Este resultado pode ser devido ao tamanho pequeno da amostra. E o resultado é o mesmo para os dois experimentos anteriores.

A Tabela 30 mostra o tempo em minutos gastos pelos participantes. Os participantes 1, 2 e 3 são os especialistas em experimentação que gastaram em média 3.9 horas para a avaliação do artigo, e os participantes 4,5 e 6 são os participantes que tinham experiência em teste, a média de tempo gasto para eles é de 8.5 horas. O artigo continha 22 páginas e 86 resultados. Pensando em termos de eficiência, os participantes no grupo 1 (especialistas em experimentação) gastaram em média de 2.94 minutos por resultado encontrado e no grupo 2 os participantes gastaram em média 10.44 minutos por resultado encontrado. Este resultado é estatisticamente significativo para o teste para diferenças entre Médias usando o método

Mann-Whitney. Estes resultados mostram que mesmo tendo gasto mais tempo avaliando o artigo, os especialistas em teste têm uma eficiência menor na extração dos resultados.

**Tabela 30 - Análise de Eficiência - Experimento III**

Participante	Especialista em:	% de Resultados Encontrados	Tempo gasto	Efetividade Média por Especialidade	Eficiência Média por Especialidade (em minutos)
S1	Exp.	65%	180	73.33%	2.94 min
S2	Exp.	90%	260		
S3	Exp.	74%	145		
S4	Teste	58%	540	58.33	10.44 min
S5	Teste	79%	840		
S6	Teste	38%	269		

## 5.6 Visão Geral dos Estudos Realizados

Para avaliar a SecESE realizamos alguns estudos de viabilidade, alguns experimentos e uma entrevista.

O primeiro estudo que fizemos, foi um estudo de viabilidade teve como objetivo a análise dos resultados de artigos relacionados com defeitos.

Depois do estudo de viabilidade, executamos um experimento controlado em Maryland no âmbito de uma disciplina de pós-graduação em Engenharia de Software Experimental. Realizamos o estudo em uma população de 7 estudantes analisando artigos técnicos em diversas áreas de engenharia de software. Neste estudo, apesar de os alunos terem coletado as informações de contexto a contento, a extração dos resultados foi abaixo da nossa expectativa inicial.

O segundo experimento foi realizado no Brasil, também em uma disciplina de pós-graduação em Engenharia de Software Experimental. Este estudo foi executado em uma população de 17 estudantes que aplicaram a técnica em artigos na área de testes de software. Este experimento foi realizado no Brasil, e tivemos vários problemas com a língua inglesa, que afetou o resultado da efetividade do experimento.

---

Após a realização do segundo experimento, fizemos uma entrevista semi-estruturada com dois dos três autores de um survey na área de teste. Esta entrevista visou capturar como os autores extraíram as informações dos artigos e comparar esta abordagem manual com a abordagem recomendada pela nossa técnica de extração. Usamos o mesmo survey para realizar um estudo de viabilidade sobre a análise de contextos experimentais utilizando técnicas de aglomeração de dados.

O último experimento do contexto deste trabalho foi realizado somente com participantes especialistas na área de experimentação ou de teste de software; o objetivo principal era analisar a influência da experiência na efetividade do uso da técnica. E vimos que os pesquisadores com experiência em experimentação têm uma eficiência maior na extração dos resultados do que os pesquisadores com experiência somente no tópico dos artigos analisados.

Estes estudos não exploram exaustivamente todas as dimensões da SecESE. Estes estudos foram projetados a partir das principais questões que surgiam à medida que o trabalho foi desenvolvido. Existem atualmente novas questões abertas a partir destes estudos. Entre elas podemos citar por exemplo: Até que ponto podemos automatizar a SecESE? Quais são as interações do processo com os processos tradicionais de revisões sistemáticas? Quais são as influências reais da qualidade dos artigos na extração das informações?

## Capítulo 6

### Conclusão

#### 6.1 Principais Resultados

Necessitamos executar estudos experimentais a fim de fornecer um melhor apoio à decisão sobre as práticas de desenvolvimento de software e seus efeitos em projetos. Devido à larga escala de fatores e ao fato de não podermos ainda especificá-los com confiança antes de realizar um estudo experimental em ES, necessitamos pensar na engenharia de software como “a grande ciência” [134].

A engenharia de software necessita executar estudos grandes ou combinar conjuntos de estudos sobre seus assuntos de interesse. Executar estudos muito grandes é uma prática que não é executada ainda na engenharia de software. As primeiras iniciativas neste sentido estão sendo discutidas somente agora no escopo da Rede Internacional da Pesquisa da Engenharia de Software (International Software Engineering Research Network - ISERN<sup>6</sup>). Entretanto, é ainda um objetivo distante buscar que a engenharia de software tenha estudos tais como aqueles realizados na física de partículas. Considere por exemplo dois artigos que anunciaram a descoberta da partícula subatômica quark. Estes artigos marcaram o fim de uma busca de 17 anos e foram realizados por 437 autores de 35 instituições e por 403 autores de 42 instituições, respectivamente [43, 117].

Uma vez que estamos longe de realizar estudos com a participação de centenas de cientistas, combinar conjuntos de estudos parece ser uma abordagem mais prática neste momento. De fato, nos últimos anos tem crescido o número de trabalhos que combinam conjuntos grandes de estudos para alcançar resultados significativos sobre processos, métodos e tecnologias. Entretanto, esta tarefa é complexa. Os estudos são feitos independentemente e não temos uma padronização de terminologia, de procedimentos de

---

<sup>6</sup> <http://isern.iese.de/network/ISERN/pub/>

---

coleta de dados e nem de apresentação dos resultados. Necessitamos de boas metodologias para extrair, combinar e analisar resultados experimentais de estudos feitos independentemente. Este é um dos desafios na área de engenharia de software experimental atualmente. Segundo Sjøberg et. al. [32], abordagens para análise secundária de estudos são uma ferramenta chave para fortalecer a Engenharia de Software como uma disciplina que está se movendo para uma maturidade completa.

Nesta tese introduzimos uma abordagem sistemática para analisar a informação dos artigos experimentais em engenharia de software - SecESE. As atividades principais da SecESE são Planejamento, Extração, Avaliação de Qualidade e Análise Secundária. Na atividade de Planejamento, os detalhes da análise são definidos e registrados na base de conhecimento; usamos uma abordagem baseada em GQM para a definição dos objetivos de análise. Na extração das informações dos artigos definimos uma técnica específica para extração das informações de artigos em ESE. Na avaliação da qualidade os estudos usados na análise são avaliados segundo os critérios de qualidade estabelecidos pelo analista e, na análise secundária, os resultados coletados nas fases anteriores são analisados; para esta atividade definimos o uso de mineração de dados no apoio desta atividade.

Os resultados experimentais e suas informações de contexto extraídos dos artigos assim como todas as informações sobre a análise, inclusive as conclusões e decisões tomadas durante a SecESE, são armazenados em uma base estruturada de resultados. A criação desta base tem diversas vantagens:

- a) A base estruturada é reusável, no sentido de que pode ser reusada para testar outras questões de pesquisa que envolvam os artigos que já estão armazenados na base;
- b) Se outro pesquisador necessitar avaliar os resultados da análise, o processo é repetível já que as informações são amplamente armazenadas na base;
- c) O processo de análise pode ser realizado por investigadores que não têm o conhecimento aprofundado sobre o contexto e os detalhes dos estudos individuais. A

---

informação estruturada ajuda o investigador a ver a informação importante para fazer a análise;

- d) Pesquisadores podem usar mineração de dados para explorar a base, como mostrado no Capítulo 4.

Para avaliar a SecESE realizamos alguns estudos de viabilidade, alguns experimentos e uma entrevista. O primeiro estudo que fizemos foi para analisar resultados de artigos relacionados com defeitos. Realizamos dois experimentos para avaliar a extração das informações de artigos de maneira sistemática. Depois focamos na avaliação da técnica de extração de informações proposta neste trabalho; executamos então um experimento controlado em Maryland em uma população de 7 estudantes analisando artigos técnicos em diversas áreas de engenharia de software. O segundo experimento controlado foi realizado no Brasil com uma população de 17 estudantes que analisaram artigos na área de testes de software que haviam sido previamente utilizados em um survey da área. Após a realização do segundo experimento, fizemos uma entrevista semi-estruturada com dois dos três autores de um survey na área de teste. Esta entrevista visou capturar como os autores extraíram as informações dos artigos e comparar esta abordagem manual com a abordagem recomendada pela técnica de extração da SecESE. Usamos o mesmo survey para realizar um estudo de viabilidade sobre a análise de contextos experimentais utilizando técnicas de aglomeração de dados. O último experimento do contexto deste trabalho foi um experimento somente com especialistas nas áreas de experimentação ou de teste de software; o objetivo principal foi analisar a influência da experiência na efetividade do uso da técnica.

A SecESE pode ser vista portanto como uma abordagem mais formal de contagem de votos que evidencia a repetibilidade do processo e formalização de atividades que não são normalmente realizadas numa abordagem de contagem de votos (como controle de qualidade dos artigos). Um das vantagens da SecESE é a formalização e garantia de qualidade da análise pela execução de um processo repetível. Além disto, a abordagem de se analisar os resultados num nível de abstração que não sejam os dados traz o benefício de inclusão, já que tanto resultados qualitativos como quantitativos podem ser usados para a análise.

---

É importante notar que muitas partes da SecESE apresentada aqui podem ser usadas junto com os outros métodos de combinar resultados de estudos. Pode ser útil por exemplo para analisar os resultados coletados para revisões sistemáticas quando a meta análise não é possível, ou talvez até para inspecionar a priori os resultados que serão explorados quantitativamente usando a meta análise.

## **6.2 Principais Contribuições**

Identificamos neste trabalho cinco contribuições principais à área de Engenharia de Software Experimental (ESE):

- i) Uma abordagem acumulativa para extração e análise de resultados experimentais em ES;
- ii) Uma estrutura para a descrição e planejamento da análise dos estudos secundários, baseada na estrutura padrão de definição de objetivos do GQM;
- iii) Uma técnica estruturada para extrair a informação de documentos científicos em ES. A abordagem SecESE define os tipos fundamentais de informações que devem ser coletadas para cada estudo e abre espaço para a inclusão de novas categorias. Isto monta um arcabouço que permite a construção do repositório mencionado no item (v);
- iv) Um método para analisar visualmente resultados e extrair conclusões baseadas no contexto e na qualidade em que estes resultados foram obtidos;
- v) A estrutura básica de um repositório que empacota parte do conhecimento implícito que surge durante o processo de análise, permitindo o reuso do repositório para análises diferentes usando a mesma base.

## **6.3 Limitações do Trabalho**

Como uma abordagem nova para a análise secundária de estudos em engenharia de software, neste trabalho focamos na definição horizontal da abordagem. Como descrito no Capítulo 3 definimos diretrizes para a execução de cada uma das atividades da SecESE como uma abordagem inicial de verticalização da definição do processo. Os estudos realizados

---

durante o trabalho foram todos no sentido de avaliar a viabilidade de aplicação da SecESE. Mas a avaliação da SecESE como um todo foi limitada; necessitamos realizar mais estudos para avaliar e aprimorar a SecESE na sua totalidade. Vislumbramos as seguintes limitações para cada macro atividade da SecESE:

- 1) Planejamento: será necessário realizar um estudo comparativo da SecESE com outras abordagens como as que são utilizadas nas revisões sistemáticas para estabelecimento das questões de pesquisa e seleção dos artigos.
- 2) Extração: será necessário estabelecer estudos adicionais que investiguem os efeitos da experiência do leitor na efetividade da extração das informações dos artigos, assim como analisar os efeitos da qualidade dos artigos. E, por meio destes estudos, aprimorar a técnica para a geração de heurísticas que poderão ser usadas na automação do processo de extração das informações.
- 3) Avaliação de Qualidade: os critérios estabelecidos neste trabalho necessitam ser refinados e aplicados a mais estudos para avaliação da sua aplicabilidade prática e o seu efeito na análise dos resultados secundários.
- 4) Análise Secundária: alguns pontos da SecESE necessitam de uma maior formalização para que possam ser estabelecidos de tal forma que seja possível a criação de ferramentas para auxiliar ainda mais o processo de análise dos resultados. Além disto, novas formas de análise podem ser incorporadas ao processo, possibilitando maior flexibilidade na escolha das técnicas de mineração.

#### **6.4 Trabalhos Futuros**

As análises secundárias estão transformando-se rapidamente na abordagem escolhida pelos pesquisadores para integrar a evidência da literatura de engenharia de software. O processo de revisão sistemática por exemplo, requer que um usuário identifique uma coleção detalhada de artigos, extraia a informação desses artigos, verifique a exatidão dos fatos extraídos, e analise os fatos extraídos usando técnicas qualitativas ou quantitativas. Embora uma revisão

---

sistemática capture exatamente a evidência, o processo é custoso, e pode demorar meses e até anos desde a concepção até a publicação [10, 57].

Conseqüentemente, é inquestionável que a área lucraria com ferramentas e métodos que poderiam ajudar a encontrar, organizar, e sumarizar informações para revisões sistemáticas, assim como sintetizá-las em conhecimento usável [26]. Uma pergunta que ainda não foi estabelecida na área é: Tais ferramentas podem ser construídas? O estabelecimento de uma abordagem como a SecESE tem como um dos objetivos o início de esforços para que num futuro próximo possamos investigar o uso da mineração do texto para realizar algumas dessas tarefas. Em particular, no uso de técnicas de extração da informações, na localização e organização de documentos para análises secundárias.

A mineração do texto busca procurar padrões em textos em linguagem natural. Ela reconhece que a compreensão completa do texto da língua natural não é atingível e foca na extração de partes pequenas de informação do texto com confiabilidade elevada [93].

A extração da informação é uma técnica usada para detectar a informação relevante em documentos maiores e apresentá-la em um formato estruturado. É usada para analisar o texto e encontrar partes específicas do texto [72]. A extração de informação é uma das técnicas mais proeminentes usadas atualmente em mineração de textos. É um ponto de partida para a análise de textos não estruturados (como os encontrados em artigos técnicos). Em particular, a combinação de ferramentas de processamento de linguagem natural, recursos léxicos, e restrições semânticas, pode fornecer os módulos eficazes para minerar documentos de vários domínios [72]. Embora esta abordagem já seja usada para revisões sistemáticas em outras áreas de conhecimento [9, 57, 72], os investigadores em ESE ainda extraem a informação dos artigos manualmente. Não há nenhuma ferramenta específica para a área, mas as técnicas de extração de informação podem ajudar significativamente a área se forem desenvolvidas ferramentas específicas para a engenharia de software.

Como trabalhos futuros em prosseguimento a esta tese, iremos investigar a adaptação de técnicas de extração de informações para ESE. Mesmo cientes de que adaptar uma técnica de extração de informações a novas áreas seja um processo custoso, que requer tanto

conhecimentos de extração de informação quanto do domínio específico da área de Engenharia de Software no qual a ferramenta será aplicada, a nossa posição é que este esforço tem muito valor tanto da perspectiva da eficiência na extração das informações como com relação à repetibilidade das extrações das informações dos artigos.

Além de investir na automatização da SecESE, pretendemos investigar maneiras de superar as limitações descritas na última seção; especialmente, pretendemos usar a SecESE em contextos maiores.

## Referências Bibliográficas

1. A. Dniestrowski, J. M. Guillaume, R. Mortier. Software Engineering in Avionics Applications. *Proceedings of the 3rd International Conference on Software Engineering*, pages 124-131, IEEE Press, 1978.
2. Adam Porter, Philip M. Johnson. Assessing Software Review Meetings: Results of a Comparative Analysis of Two Experimental Studies. *IEEE Transactions on Software Engineering*. 23(3):129-145, March, 1997.
3. Adam Porter, Richard W. Selby. Empirically Guided Software Development Using Metric-Based Classification Trees. *IEEE Software*, 7(2): 46-54, Mar/Apr, 1990.
4. Alan M. Davis, Óscar Dieste Tubío, Ann M. Hickey, Natalia Juristo Juzgado, Ana María Moreno: Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review. *RE 2006*, 176-185, 2006.
5. Albert Endres. An analysis of Errors and their Causes in System Programs, *ACM SIGPLAN Notices*, 10(6): 327-336, Los Angeles, CA, April 1975.
6. Andreas Zendler. A Preliminary Software Engineering Theory as Investigated by Published Experiments. *Empirical Software Engineering*, 6(2):161-180, June, 2001.
7. Andy Brooks. Meta Analysis-A Silver Bullet-for Meta-Analysts. *Empirical Software Engineering Journal*, 2(4): 333-338, 1997.
8. Anil K. Jain, M. Narasimha Murty, Patrick J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264-323, September, 1999.
9. Anna Divoli , Teresa K. Attwood. BioIE: Extracting Informative Sentences from the Biomedical Literature, *Bioinformatics*, 21(9): 2138-2139, May, 2005.
10. Anthony Petrosino. *Lead Authors of Cochrane Reviews: Survey Results*. Report to the Campbell Collaboration, University of Pennsylvania, Cambridge, MA, 1999.
11. Australian National Health and Medical Research Council. *How to review the evidence: systematic identification and review of the scientific literature*, 2000. IBSN 186-4960329.
12. Australian National Health and Medical Research Council. *How to use the evidence: assessment and application of scientific evidence*. February 2000, ISBN 0642432952.
13. Barbara A. Kitchenham, Emilia Mendes, Guilherme H. Travassos. Cross versus Within-Company Cost Estimation Studies: A Systematic Review. *IEEE Transactions on Software Engineering*, 33(5): 316-329, May, 2007.

14. Barbara Kitchenham, Emilia Mendes, Guilherme Horta Travassos. Systematic Review of Cross- vs. Within- Company Cost Estimation Studies. Proceedings of EASE 2006: *Evaluation & Assessment in Software Engineering*, BCS-eWIC, pages 89–98, 2006.
15. Barbara Kitchenham. Evaluating software engineering methods and tools, *ACM SIGSOFT Software Engineering Notes*, 21(1):11-15, 1996.
16. Barbara Kitchenham. Procedures for undertaking systematic reviews. Technical Report TR/SE-0401, Department of Computer Science, Keele University and National ICT, Australia, 2004.
17. Barry Boehm. *Software Engineering Economics*. Prentice-Hall, 1981.
18. Ben Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics (TOG)*, 11(1): 92-99, 1992.
19. Bernd Freimut, Teade Punter, Stefan Biffel, Marcus Ciolkowski. State-of-the-Art in Empirical Studies. ViSEK Technical Report 007/E, 2002.
20. Brian Johnson, Ben Shneiderman. *Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures*. Proc. of IEEE Information Visualization '91, pages 284 - 291, 1991.
21. Brian Marick, A Survey of Software Fault Surveys, Technical Report UIUCDCS-R-90-1651, University of Illinois, 1990.
22. Carla M. Dal Sasso Freitas, Paulo R. G. Luzzardi, Ricardo A. Cava, Marco A. A. Winckler, Marcelo S. Pimenta, Luciana P. Nedel. Evaluating Usability of Information Visualization Techniques. *5th Symposium on Human Factors in Computer Systems (IHC)*, Brazilian Computer Society Press, Fortaleza, 2002.
23. Carla Maria Dal Sasso Freitas, Olinda Mioka Chubachi, Paulo Roberto Gomes Luzzardi, Ricardo Andrade Cava. Introdução à visualização de informações. *Revista de Informática Teórica e Aplicada*, Porto Alegre, 8(2):143-158, Outubro, 2001.
24. Carol Withrow, Error density and size in Ada software, *IEEE Software*, 7(1):26-30, January, 1990.
25. Carolyn B. Seaman. Qualitative Methods in Empirical Studies of Software Engineering. *IEEE Transactions on Software Engineering*, 25(4):557-572, Aug/Sept. 1999.
26. Catherine Blake. Information Synthesis: A New Approach to Explore Secondary Information in Scientific Literature. *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, pp 56-64, Denver, CO, US, 2005.

- 
27. Christiane da Costa Santana. Uma Ferramenta para Construção interativa de Árvores de Classificação Usando Mapas em Árvore. Dissertação de Mestrado - Universidade Salvador, 2004.
  28. Christiane Santana, Manoel Mendonça, Daniela S. Cruzes. Visualizing the Construction of Decision Trees Using *Treemaps*. In: *International Workshop on Visual Languages and Computing (VLC'2004)*, 1: 213-218, San Francisco, 2004.
  29. Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, Anders Wesslén. A. *Experimentation in Software Engineering – An Introduction*. Kluwer Academic Press, 2000.
  30. *Cochrane Collaboration*. *Cochrane Reviewers' Handbook*. Version 4.2.6. September 2006.
  31. D. Potier, J. L. Albin, R. Ferreol, A. Bilodeau. Experiments with Computer Software Complexity and Reliability. *Proceedings of the 6th International Conference on Software Engineering*, pages.94-103, IEEE Press, Tokyo, Japan, 1982.
  32. Dag I.K. Sjoberg, Jo E. Hannay, Ove Hansen, Vigdis By Kampenes, Amela Karahasanovic, Nils-Kristian Liborg, Anette C. Rekdal. A survey of controlled experiments in software engineering. *IEEE Trans. Software Engineering*, 31 (9): 733-753, September. 2005.
  33. Daniel A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1-8, January, 2002.
  34. Daniela S. Cruzes, Manoel G. Mendonça Neto, Christiane Santana, José C. Maldonado, Mario Jino. Construction of Predictive Software Engineering Models Using *Treemaps* Decisions Trees. *Software Technology and Engineering Practice - STEP*, Chicago. Ottawa, p. 1-4, 2004.
  35. Daniela S. Cruzes, Manoel G. Mendonça Neto, Christiane Santana. Aquisição de Conhecimento Durante a Mineração de Dados: Um Estudo de Caso em Tarefa de Classificação. *Jornada Ibero-americana de Engenharia de Software e Engenharia de Conhecimento*, Madri. Espanha: Serviço de Publicaciones de la Facultad de Informática de la UPM, v. 2. p. 677-680, 2004.
  36. Daniela S. Cruzes, Manoel G. Mendonça Neto, José C. Maldonado, Mario Jino. Using Visualization to Bring Context Information to Software Engineering Model Building. In: *International Workshop on Visual Languages and Computing (VLC'2004)*, San Francisco. *Proceedings of the 2004 International Conference on Distributed Multimedia Systems*. Skokie, Illinois: Knowledge Systems Institute, v. 1. p. 219-224, 2004.
  37. Daniela S. Cruzes, Manoel G. Mendonça Neto, Victor Robert Basili; Forrest Shull;, Mario Jino. Using Context Distance Measurement to Analyze Results across Studies. In:

- International Symposium on Empirical Software Engineering and Measurement (ESEM), 2007, Madri, Espanha. Proceedings of the 1st ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2007. v. 1. p. 1-10.
38. Daniela S. Cruzes, Manoel G. Mendonça Neto, Victor Robert Basili; Forrest Shull;, Mario Jino. Automated Information Extraction from Empirical Software Engineering Literature: Is that possible?. In: International Symposium on Empirical Software Engineering and Measurement (ESEM), 2007, Madrid. Proceedings of the 1st ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2007. v. 1. p. 1-3.
39. David M. Weiss, Victor R. Basili. Evaluating Software Development by Analysis of Changes: The Data from the Software Engineering Laboratory, *IEEE Transactions on Software Engineering*, SE-11 (2): 157-168, February, 1985.
40. Dieter H. Rombach, Victor R. Basili. A Quantitative Assessment of Software Maintenance: An Industrial Case Study. *Conference on Software Maintenance*, Austin, Texas, September, 1987.
41. Donald T. Campbell, Julian C. Stanley. *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally, 1966.
42. Fabrice Guillet and Howard J. Hamilton. *Quality Measures in Data Mining (Studies in Computational Intelligence)*. Springer-Verlag New York, Inc, 2007.
43. F. Abe et. al. Observation of Top Quark Production in  $p p$  Collisions with the Collider Detector at Fermilab. *Phys. Rev. Letter*. 74 (14): 2626-2631, 1995.
44. Forrest Shull, Daniela Cruzes, Victor Basili, Manoel G. Mendonça. Simulating Families of Studies to Build Confidence in Defect Hypotheses. *Information and Software Technology*, 47(15):1019-1032, December, 2005.
45. Forrest Shull, Manoel G. Mendonça, Victor R. Basili, Jeffrey Carver, José Carlos Maldonado, Sandra Camargo Pinto Ferraz Fabbri, Guilherme Horta Travassos, Maria Cristina Ferreira de Oliveira: Knowledge-Sharing Issues in Experimental Software Engineering. *Empirical Software Engineering*, 9(1-2):111-137, 2004.
46. Forrest Shull, Victor R. Basili, Jeffrey Carver, José Carlos Maldonado, Guilherme Horta Travassos, Manoel G. Mendonça, Sandra Camargo Pinto Ferraz Fabbri: Replicating Software Engineering Experiments: Addressing the Tacit Knowledge Problem. *ISESE 2002*, pages 7-16, 2002
47. Frederic M. Wolf. *Meta analysis—Quantitative Methods for Research Synthesis*. Sage Publications, 1986.

- 
48. G. W. Snedecor, W. G. Cochran, *Statistical Methods*, Iowa State University Press, 8th Edition, 1989, ISBN 0-813-81561-4.
  49. Gene V. Glass. Primary, secondary and meta-analysis of research. *Educational Researcher*, 5(10): 3-8, Novembro, 1976.
  50. George G. Robertson, Jock D. Mackinlay, Stuart K. Card. Cone Trees: Animated 3D Visualizations of Hierarchical Information. *Human Factors in Computing Systems (CHI '91)*. Conference Proceedings, ACM Press, pages. 189-194, 1991.
  51. Gregg B. Jackson. Methods for integrative reviews, *Review of Educational Research*, 50(3): 438-460, 1980.
  52. Gregory Piatetsky-Shapiro, Christopher J. Matheus. The interestingness of deviations. *In Proceedings of AAAI-9, Workshop on knowledge Discovery in Databases*, pages 25-36, 1994
  53. Guilherme H. Travassos, Dmytro Gurov, Edgar A. G. do Amaral. Introdução à Engenharia de Software Experimental. Relatório Técnico ES-590/02-Abril, Programa de Engenharia de Sistemas e Computação, COPPE/UFRJ. 2002.
  54. Guilherme H. Travassos, Marcio O. Barros. Contributions of In Virtuo and In Silico Experiments for the Future of Empirical Studies in Software Engineering. Proc. of *2nd Workshop in Workshop Series on Empirical Software Engineering the Future of Empirical Studies in Software Engineering*, Italy, 2003.
  55. Haim Levkowitz. Color icons: Merging color and texture perception for integrated visualization of multiple parameters. *In Proc. Visualization 91*, San Diego, CA, pages. 164-170, 1991
  56. Harris M. Cooper. Scientific Guidelines for Conducting Integrative Research Reviews. *Review of Educational Research*, 52(2): 291-302, 1982.
  57. Hisham Al-Mubaid. A Text-Mining Technique for Literature Profiling and Information Extraction from Biomedical Literature, *ISSO Y2005 Annual Report* , p. 45-49, 2005.
  58. Ian H. Witten, Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, 2000.
  59. James Miller. Applying meta-analytical procedures to software engineering experiments, *Journal of Systems and Software*, 54(1): 29-39, September, 2000.
  60. James Miller. Replicating software engineering experiments: a poisoned chalice or the Holy Grail. *Information & Software Technology*, 47(4): 233-244, 2005.
  61. James Warner. Clinicians' Guide to reading psychiatric literature: therapeutic trials and systematic reviews. *Advances in Psychiatric Treatment*, 8, 73-80, 2002.

62. Jinwook Seo. Information Visualization Design for Multidimensional Data: Integrating the Rank-By-Feature Framework with Hierarchical Clustering. Ph.D. Dissertation, Dept. of Computer Science, University of Maryland, December, 2005.
63. Jo E. Hannay, Dag I.K. Sjøberg, Tore Dybå, A Systematic Review of Theory Use in Software Engineering Experiments. *IEEE Transactions on Software Engineering*, 33(2): 87-107, February, 2007.
64. John Daly. Replication and a Multi-Method Approach to Empirical Software Engineering Research. PhD Thesis, Department of Computer Science, University of Strathclyde, 1996.
65. Jorge Biolchini, Paula G. Mian, Ana C. Natali, Guilherme H. Travassos. Systematic Review in Software Engineering: Relevance and Utility. Technical Report ES679/05, PESC - COPPE/UFRJ, 2005. <http://cronos.cos.ufrj.br/publicacoes/reltec/es67905.pdf>.
66. Jorge Calmon de Almeida Biolchini, Paula Gomes Mian, Ana Candida Cruz Natali, Tayana Uchôa Conte, Guilherme Horta Travassos. Scientific research ontology to support systematic review in Software Engineering. *Advanced Engineering Informatics*, 21(2):133-151, April, 2007.
67. José C. Maldonado, Jeffrey Carver, Forrest Shull, Sandra Fabbri, Emerson Dória, Luciana Martimiano, Manoel Mendonça, Victor Basili. Perspective-Based Reading: A Replicated Experiment Focused on Individual Reviewer Effectiveness, *Empirical Software Engineering*, 11(1) 119-142, March, 2006.
68. José C. Maldonado, Luciana Martimiano, Emerson Dória, Sandra Fabbri, Manoel Mendonça. Readers Project: Replication of Experiments A Case Study using Requirements Documents. In: Workshop de Avaliação e Mostra dos resultados dos projetos cooperativos de Pesquisa, Desenvolvimento e Inovação em Tecnologia da Informação e Comunicação, 2001, Rio de Janeiro. Anais do Workshop de Avaliação e Mostra dos resultados dos projetos cooperativos de Pesquisa, Desenvolvimento e Inovação em Tecnologia da Informação e Comunicação, pages 85-117, 2001.
69. José Carlos Maldonado, Jeffrey Carver, Forrest Shull, Sandra Fabbri, Emerson Dória, Luciana Martimiano, Manoel G. Mendonça, Victor R. Basili: Perspective-Based Reading: A Replicated Experiment Focused on Individual Reviewer Effectiveness. *Empirical Software Engineering*, 11(1): 119-142, 2006.
70. Josemeire Machado Dias. Utilizando Estudos Observacionais para Testar e Aperfeiçoar um Pacote de Laboratório para Avaliação de Ferramentas de Mineração Visual de Dados. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador, 2006.

- 
71. Karina M. Villela. Ambientes de Desenvolvimento de Software Orientados a Organização. Tese de Doutorado em Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro, 2004.
  72. Katharina Kaiser and Silvia Miksch. Information Extraction: A Survey. Vienna University of Technology, Institute of Software Technology and Interactive Systems, Vienna, Technical Report, Asgaard-TR-2005-6, May 2005.
  73. Kevin Strike, Khaled El Emam, Nazim Madhavji. Software Cost Estimation with Incomplete Data. *IEEE Transactions on Software Engineering*, 27(10): 890-908, Oct., 2001.
  74. Khalid S Khan, Gerben ter Riet, Julie Glanville, Amanda J Sowden and Jos Kleijnen (editors). Undertaking Systematic Review of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2<sup>nd</sup> Edition), NHS Centre for Reviews and Dissemination, University of York, March 2001.
  75. Lesley Pickard, Barbara A. Kitchenham, Peter Jones. Combining Empirical Results in Software Engineering. *Information and Software Technology*, 40(14): 811-821, 2000.
  76. Lionel C. Briand, Victor R. Basili, William M. Thomas. A Pattern Recognition Approach for Software Engineering Data Analysis. *IEEE Transactions on Software Engineering*, 18(11): 931-942, Novembro, 1992.
  77. Lorin Hochstein. Development of an Empirical Approach to Building Domain-Specific Knowledge Applied to High-End Computing, University of Maryland, College Park, 2006.
  78. Lorrie R. Gay, Peter W. Airasian. *Educational Research: Competencies for Analysis and Application*, Fifth Edition, Prentice Hall, 1996.
  79. Luis da Camara Cascudo. *Dicionário Brasileiro de Folclore*, Edições de Ouro, 1962.
  80. Madhukar Pai, Michael McCulloch, Jennifer D. Gorman, Nitika Pai, Wayne Enanoria, Gail Kennedy, Prathap Tharyan, John M. Colford, Jr, Systematic reviews and meta-analyses: an illustrated, step-by-step guide. *The National Medical Journal of India*. 17(2): 86-95, March-April, 2004.
  81. Magne Jørgensen, Dag Sjøberg. Generalization and Theory-Building in Software Engineering Research. *Empirical Assessment in Software Engineering (EASE2004)*. IEE Proceedings, pages 29–36, 2004.
  82. Magne Jørgensen, Martin Shepperd. A Systematic Review of Software Development Cost Estimation Studies. *IEEE Transactions on Software Engineering*, 33(1):33-53, January, 2007.

83. Magne Jørgensen, Martin Shepperd. A Systematic Review of Software Development Cost Estimation Studies. *IEEE Transactions on Software Engineering*, 33(1): 33-53, January, 2007.
84. Magne Jørgensen. Experience With the Accuracy of Software Maintenance Task Effort Prediction Models. *IEEE Transactions on Software Engineering*, 21(8): 674-681, August, 1995.
85. Manoel C. Marques Neto, Manoel Mendonça, Celso A. Saibel Santos. GraphMiner: Uma Ferramenta de Mineração Visual de Dados em Bases Relacionais In: *JISIC'04, 2004, Madrid. Actas de las IV Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*. 1:305-316, Madrid,2004
86. Manoel G. Mendonça, Daniela S. Cruzes, Josemeire M. Dias, Maria C. F. Oliveira,. Using Observational Pilot Studies to Test and Improve Lab Packages. *Proceedings of the 5th ACM-IEEE International Symposium on Empirical Software Engineering*, 1: 1-10, Rio de Janeiro, 2006.
87. Manoel G. Mendonça; Daniela S. Cruzes, Christiane Santana. Interactive Construction of Classification Trees Using *Treemaps*. In: *Conferencia Latinoamericana de Informática*, Arequipa, Peru, v1. pp 1-10, 2004.
88. Manoel Mendonça, Nancy L. Sunderhaft. Mining Software Engineering Data: A Survey. A DACS State-of-the-Art Report, September, 1999. <http://www.dacs.dtic.mil/techs/datamining/>
89. Marcus Ciolkowski, Forrest Shull, Stefan Biffl. A Family of Experiments to Investigate the Influence of Context on the Effect of Inspection Techniques. In *Proceedings of the 6th International Conference on Empirical Assessment in Software Engineering (EASE)*, pages 48-60, Keele, UK, April, 2002.
90. Maria Cristina Ferreira de Oliveira, Haim Levkowitz. From Visual Data Exploration to Visual Data Mining: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 9 (3): 378-394, Jul-Sept, 2003
91. Mariano Montoni. Aquisição de Conhecimento: uma Aplicação no Processo de Desenvolvimento de Software, Tese de Mestrado, COPPE/UFRJ Agosto 2003.
92. Mark Staples, Mahmood Niazi. Experiences using Systematic Review Guidelines. *Proceedings of EASE 2006: Evaluation & Assessment in Software Engineering*, pages 79-88, 2006.
93. Marti A. Hearst. Untangling Text Data Mining, Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20-26, 1999.

- 
94. Marvin V. Zelkowitz, Dolores Wallace. Experimental models for validating Technology. *IEEE Computer*, 31(5): 23-31, May, 1998.
  95. Mary Dixon-Woods, Shona Agarwal, David Jones, Bridget Young, Alex Sutton. Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Services Research and Policy*, 10(1): 45-53, January 2005.
  96. Mauricio A. de Almeida, Hakim Lounis, Walcélío L. Melo. An Investigation on the Use of Machine Learned Models for Estimating Correction Costs. *20th International Conference on Software Engineering (ICSE'98)*, page 473, 1998.
  97. Mihael Ankerst. Visual Data Mining, Ph.D. thesis, University of Munich, published by [www.dissertation.de](http://www.dissertation.de), 2000.
  98. Murray Wood, John W. Daly, James Miller, Marc Roper. Multi-Method Research: An Empirical Investigation of Object-Oriented Technology. *Journal of Systems and Software*, 48(1): 13-26, 1999.
  99. Myron Lipow: Prediction of software failures. *Journal of Systems and Software*, 1: 71-75 July, 1979.
  100. Natalia Juristo, Ana M. Moreno, Sira Vegas. Reviewing 25 years of testing technique experiments. *Empirical Software Engineering Journal*, 9 (1-2): 7-44, March, 2004.
  101. Natalia Juristo, Ana Moreno. An Adaptation of Experimental Design to Empirical Validation of Software Engineering Theories. *23rd Annual NASA Software Engineering Workshop*, Maryland (USA): 1998.
  102. Natalia Juristo, Ana Moreno. *Basics of Software Engineerig Experimentation*, Kluwer, 2001.
  103. Nikhil Sharma. The Origin of the Data Information Knowledge Wisdom Hierarchy. December, 2005. [http://www-personal.si.umich.edu/~nsharma/dikw\\_origin.htm](http://www-personal.si.umich.edu/~nsharma/dikw_origin.htm).
  104. Oliver Niggemann. Visual Data Mining of Graph-Based Data, Ph.D. Thesis, University of Paderborn, Germany, 2001.
  105. Parastoo Mohagheghi, Reidar Conradi. Vote-Counting for Combining Quantitative Evidence from Empirical Studies - An Example. *Proc. 5th ACM-IEEE Int'l Symposium on Empirical Software Engineering (ISESE'06)*, Rio de Janeiro, IEEE Press, p. 24-26, September 2006.
  106. Paula Mian, Ana Candida Cruz Natali, Jorge Biolchini, Guilherme H. Travassos, Tayana Uchoa Conte. A Systematic Review Process for Software Engineering. *In: 3rd ESELAW - Experimental Software Engineering Latin American Workshop*, Uberlandia, 2005.

107. Paula Mian, Tayana Uchoa Conte, Ana Candida Cruz Natali, Jorge Biolchini, Guilherme Horta Travassos. Lessons Learned on Applying Systematic Reviews to Software Engineering. *In: WSESE2005 - Workshop Series in Empirical Software Engineering, 2005, Oulu. Proceedings of the 3rd International Workshop GUIDELINES FOR EMPIRICAL WORK in the Workshop Series on Empirical Software Engineering 2005.* Kaiserslautern : Fraunhofer Center, v.1, pages 1-6, 2005.
108. Pepjijn R. S. Visser, Dean M. Jones, Bench T. J. M. Capon, M. J. R. Shave. An analysis of ontological mismatches: heterogeneity versus interoperability. *In: AAAI 1997 spring symposium on ontological engineering, Stanford, pages 164-172, USA, 1997.*
109. Per Runeson, Thomas Thelin. Prospects and Limitations for Cross-study Analyses - A Study on an Experiment Series, *2nd Workshop on The Future of Empirical Studies in Software Engineering*, pages. 133-142, Rome, Italy, September, 2003.
110. R. Pickett and G. Grinstein, Iconographic displays for visualizing multidimensional data. *In Proc. IEEE Conf. on Systems, Man and Cybernetics, IEEE Press, Piscataway, NJ, pages 514-519, 1988*
111. Raymond J. Rubey. Quantitative Aspects of Software Validation. *Proceedings of the 1975 International Conference on Reliable Software, in SIGPLAN Notices*, 10 (6): 246-251, June, 1975.
112. Richard W. Selby, Victor R. Basili, Error localization during software maintenance: Generating hierarchical system descriptions from the source code alone. *In Proceedings of the Conference on Software Maintenance*, Phoenix, AZ, October 1988.
113. Richard W. Selby, Victor R. Basili. Analyzing Error Prone System Structure, *IEEE Transactions on Software Engineering*, 17(2): 141-152, February 1991.
114. Rini van Solingen, Egon Berghou. *The Goal Question Metric Method: A Practical Guide for Quality Improvement of Software Development.* McGraw-Hill, 1999.
115. Robert L. Glass, Persistent Software Errors. *IEEE Transactions on Software Engineering*, 7(2): 162-168, March, 1981.
116. Russell .L. Ackoff. From Data to Wisdom. *Journal of Applied Systems Analysis*, 16(1): 3-9, 1989.
117. S. Abachi et. al. Observation of the Top Quark. *Phys. Rev. Letter*, 74(14): 2632-2637, 1995.
118. S. Djoki, Giancarlo Succi, Witold Pedrycz, Martin P. Mintchev. Meta Analysis - A Method of Combining Empirical Results and its Application in Object-Oriented Software Systems. *7th International Conference on Object Oriented Information System (OOIS'2001)*, pages 103-112, 2001.

- 
119. Sima Asgari, Lorin Hochstein, Victor Basili, Jeff Carver, Jeff Hollingsworth, Forrest Shull, Generating Testable Hypotheses from Tacit Knowledge for High Productivity Computing. *In Proceedings of the Workshop on Software Engineering and High Performance Computing Applications (held at ICSE 2005)*. St. Louis, MO, USA, 2005.
  120. Stuart K. Card , Jock D. Mackinlay , Ben Shneiderman. Information Visualization. Readings in Information Visualization -Using Visualization to Think. San Francisco, Morgan Kaufmann Publ, pages pp. 1-34, 1999.
  121. Stuart K. Card , Jock D. Mackinlay. The structure of information visualization design space. *Proceedings of the Information Visualization Symposium*, Phoenix, Arizona. IEEE Computer Society, pages 92-99, 1997.
  122. T. E. Bell and T. A. Thayer. Software Requirements: Are They Really a Problem? *Proceedings of the 2nd International Conference on Software Engineering*, pages 61-68, IEEE Press, 1976.
  123. Thomas D. Cook, Harris Cooper. *Meta-Analysis for Explanation*, New York, Russell Sage Foundation, 1992
  124. Thomas J. Ostrand, Elaine J. Weyuker. Collecting and Categorizing Software Error Data in an Industrial Environment. *Journal of Systems and Software*, 4 (4):289-300, November, 1984.
  125. Tore Dyba, Barbara A. Kitchenham, Magne Jorgensen. Evidence-Based Software Engineering for Practitioners. *IEEE Software*, 22(1): 58-65, Jan/Feb, 2005.
  126. Tore Dybå, Vigdis By Kampenes, Dag I. K. Sjøberg: A systematic review of statistical power in software engineering experiments. *Information & Software Technology*, 48(8): 745-755, 2006.
  127. Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11): 27-34, November 1996.
  128. Usama Fayyad, Ramasamy Uthurusamy. Data mining and knowledge discovery in databases. *Communications of the ACM*, 39(11):24-26, November 1996.
  129. Victor R. Basili, Barry T. Perricone. Software Errors and Complexity: An Empirical Investigation, *Communication of the ACM*, 27(1): 42-52, January, 1984.
  130. Victor R. Basili, David M. Weiss. Evaluation of Requirements Document by Analysis of Change Date, *Proceedings of the Fifth International Conference on Software Engineering*, pages 314-323, March 1981.

131. Victor R. Basili, Dieter Rombach, Richard W. Selby. The Experimental Paradigm in Software Engineering. *Experimental Software Engineering Issues: Critical Assessment and Future Directions, International Workshop*, Dagstuhl, Germany, 1992. Appeared in Springer-Verlag, Lecture Notes in Computer Science, 706, 1993.
132. Victor R. Basili, Forrest Shull, Filippo Lanubile. Building Knowledge through Families of Experiments. *IEEE Transactions on Software Engineering*, 25(4): 456-473, Julho, 1999.
133. Victor R. Basili, Gianluigi Caldiera, Dieter Rombach. Goal Question Metric Approach. *Encyclopaedia of Software Engineering*. pp. 528-532, John Wiley & Sons, 1994.
134. Victor R. Basili. The Past, Present, and Future of Experimental Software Engineering. *Journal of the Brazilian Computer Society (JBACS)*, pp 7-12, December, 2006.
135. Victor R. Basili. The Role of Experimentation: Past, Current, and Future. *18th International Conference on Software Engineering*, pages. 442-450, 1996.
136. Vincent Yun Shen, Tze-Jie Yu, Stephen M. Thebaut, Lorri R. Paulsen. Identifying Error-Prone Software - An Empirical Study, *IEEE Transactions on Software Engineering*, 11(4):317-324, April, 1985.
137. Will Hayes. Research Synthesis in Software Engineering: A Case for Meta-Analysis. *Sixth International Software Metrics Symposium (METRICS'99)*, pages 143-151, 1999.
138. Yasuhiro Mashiko, Victor R. Basili. Using the GQM Paradigm to Investigate Influential Factors for Software Process Improvement, *Journal of Systems and Software*, 36(1): 17-32, January, 1997.

# **Apêndice A**

## **Formulários dos Experimentos**

# Background Form

Date

\_\_/\_\_/\_\_

This form asks you a few questions about your background and experience. Please fill and send to daniela@unifacs.br.

Thank you!

## General Background

<b>Name</b>	
<b>E-mail</b>	
<b>Number of Years of Experience in Software Engineering</b>	

## Empirical Studies Experience

Please rate your experience in the following activities:

For Experience Level use the following scale.

0 = No experience (Leave Extra Information blank)

1 = Classroom experience only

On the Extra information column indicate whether you:

a) Only learned the concept in class

b) Used the concept on a homework

c) Used the concept on a project

2 =Professional experience

Please indicate on the Extra information column the number of projects on which you performed the activity

	Experience on Empirical Studies	
	Experience Level	Extra Information
<b>Participating as a Subject</b>		
<b>Replicating the Experiment</b>		
<b>Designing the Experiment</b>		

# Consent Form

Date

\_\_/\_\_/\_\_

- **Título do Projeto e Objetivo**

O Experimento é chamado de "Experimento para avaliação de Técnicas de Extração de Informações de Relatórios". Para este experimento em particular, a técnica chamada "Information Extraction Technique".

- **Objetivos**

Proposta:	<b>Analisar</b>	Os formulários de coleta de informações
	<b>Com o propósito de</b>	Avaliar uma técnica de extração de informações de artigos.
Perspectiva:	<b>Com respeito</b>	À consistência e completude das informações extraídas
	<b>Do ponto de vista de</b>	Estudantes do Curso de Engenharia de Software Experimental

- **Procedimentos**

O experimento envolve duas sessões, uma para Treinamento com aproximadamente 2 horas e outra para aplicação da técnica em dois artigos, cada uma com um tempo estimado de 2 horas em casa, com a seguinte programação:

05/09 Treinamento – durante a aula

12/09 Entrega dos Resultados da Aplicação da Técnica em dois artigos

- **Confidencial**

Todas as informações coletadas no experimento são confidenciais e meu nome não será identificado em momento algum. Confirmando que sou maior de 18 anos e que desejo participar do experimento conduzido pelo professor Dr Manoel Mendonça e a Professora Daniela Cruzes.

Eu entendo que não serei remunerado pessoalmente por participar do experimento, porém a minha participação me ajudará a aprender mais sobre experimentação em Engenharia de Software. Eu compreendo que ficarei livre para fazer perguntas ou para deixar de participar do experimento a qualquer momento sem sofrer penalidades.

Eu, \_\_\_\_\_ aceito participar do experimento de acordo com o descrito acima.

# Post Study Questionnaire

Date

\_\_/\_\_/\_\_

These questions are to help us improve the experiment and interpret the results. The answers will not affect your grade in anyway. Please fill out and send to daniela@unifacs.br by Monday, October 9<sup>th</sup>.

Thank you!

<b>Name:</b>	
	Rate your English Level:  <input type="checkbox"/> Low  <input type="checkbox"/> Medium  <input type="checkbox"/> High

## Impressions about the Technique

Please, indicate your agreement with each statement below then justify your answers:

1) The Technique is useful for extracting information from papers.

                                                                                         
Totally Disagree    Somewhat Disagree    No Opinion    Somewhat Agree    Totally Agree

---

---

---

---

2) If I had more training I would have be able to apply it better.

                                                                                         
Totally Disagree    Somewhat Disagree    No Opinion    Somewhat Agree    Totally Agree

---

---

---

---

## Impressions about the Results

Please, indicate your agreement with each statement below then justify your answers:

---

3) Using the technique I was able to extract all important Hypotheses from the paper.

Totally Disagree    Somewhat Disagree    No Opinion    Somewhat Agree    Totally Agree

4) The fields on the context information captured most of what I felt was important about the experimental context.

Totally Disagree    Somewhat Disagree    No Opinion    Somewhat Agree    Totally Agree

5.a What fields were missing and which did you think were not relevant?

---

---

---

5) If I had more time, I would have been able to apply the technique better.

Totally Disagree    Somewhat Disagree    No Opinion    Somewhat Agree    Totally Agree

6) If I was going to generalize results across multiple papers, I could only need to look at the extracted information and get the information I needed (that is, I wouldn't have to go back and read the papers again).

Totally Disagree    Somewhat Disagree    No Opinion    Somewhat Agree    Totally Agree

---

### Suggestions

Please, give suggestions about how we can improve our work.

---

7) In your opinion, how can we improve the technique?

---

---

---

---

8) In your opinion, how can we improve the experiment?

---

---

---

---

# Information Extraction Technique

This document describes the instructions for using the Information Extraction Technique.

The document is divided into 5 main sections:

1. An introduction to the technique
2. A glossary where the key terms are defined for this context
3. A procedure of steps that describes clearly WHAT you have to do.
4. A set of guidelines, i.e. heuristics to keep in mind as you do the above procedure.
5. A set of data entry forms

## 1) Introduction

The Information Extraction Technique is a structured reading method for extracting information from papers, which can later be analyzed to explore the evidence that supports various hypotheses. In this assignment, you will follow the procedure for identifying and recording results from papers in the scientific literature. The procedure focuses the search specifically on results and context descriptions, providing some guidelines to help recognize and abstract them.

## 2) Glossary

**Result:** A result is a tentative explanation for certain behaviors, phenomena, or events that have occurred. A good result statement states as clearly and concisely as possible the relationship (or difference) between two or more variables and defines those variables in operational, measurable terms. For the purpose of our analysis, we classify results as tested or “from other papers”:

### a) Tested Results

A tested result is a tentative explanation for certain behaviors, phenomena, or events that have occurred, based upon experience or empirical study in this paper.

### b) Results of Empirical Studies from another paper, from experience

A result tested in another paper and reported in this one.

## 3) Procedure

- 1) Read the paper, keeping in mind the two kinds of information that you want to identify:

- a. Results (tested and from another paper), and
  - b. Context descriptions
- 2) When you find relevant information during your reading, highlight it so that there can be some traceability back to the original source if questions arise later.
  - 3) Transfer the key details to the data entry forms. For complete descriptions of the fields you should complete, see section 5 on “data entry forms.”

#### 4) Some Guidelines

Papers are implicitly broken up into several sections: an introduction, the method used, the analysis of the results, and the interpretation of the results.

The Introduction of the paper usually sets the research in a context (it provides the "big picture"), provides a review of related research, and develops the hypotheses for the research, these hypotheses will turn out to be results in the analysis section.

The “Method” section is usually a description of how the research was conducted, including who the participants were, the design of the study, what the participants did, and what measures were used. This section will probably contain the information to be filled on the context description template. There should be at least one context description associated with each source (possibly more, if the paper describes data that was collected from several projects). Our experience is that the context descriptions usually come shortly after the introduction. As different studies report different metrics of interest to them, not every paper will have all of the required information for our template. However, the template should be filled out as completely as possible given the information that has been published in the paper.

An “Analysis of Results” section usually describes the outcomes of the measures of the study. A Discussion section contains the interpretations and implications of the study. There may be more than one study in the report; in this case, there are usually separate Method and Results sections for each study followed by a General Discussion or sometimes a Conclusion that ties all the research together. Our experience is that these are the sections of the paper where most tested hypotheses can be found. The conclusions are a good place to find the main results, although these are many times repetitions from earlier in the text.

The results of the study are used by the researchers to answer the research questions through summaries and analyses of the measures obtained in the study, which can usually be found when reading the Analysis of Results, Discussion and Conclusion sections. (Is this what you meant?) When you identify a statement that you believe to be a result, use the following questions to help you confirm that it is a result:

- Does this statement
  - State results of measurements?
  - summarize the raw data?
  - highlight some specific characteristic of the raw data?
  - Provide insights about tables and figures?

- Summarize the results of statistical analyses?
- can be used to answer the research question(s)?
- reflects the main results of the study?

If the answer is YES to any of these questions, the statement should be collected as a Result. It's important to notice that the most important results to be gathered are the results that can be generalized beyond the context of the study.

Some results will also be found in tables and figures; although not explicitly stated in the text of the paper, relationships that are expressed visually for readers will need to be translated into textual form to be inserted into the Results Form in a usable way.

A report of an empirical study also includes an Abstract that provides a very brief summary of the research and a References section that contains information about all the articles and books that were cited in the report.

Here are some statistics we gathered about the collection of results from 22 papers:

	Tested Results	Results From Other
Total of Results Analyzed	164	72
<b>Discussion and Conclusions</b>	<b>32.93%</b>	5.56%
<b>Experimental Results</b>	<b>56.10%</b>	15.28%
<b>Introduction</b>		<b>20.83%</b>
<b>Method</b>	9.15%	8.33%
<b>Related Work</b>		<b>50.00%</b>

Note that if you focus on Discussion and Conclusion and Experimental Results sections you should find 89% of the hypotheses. With regard to identifying results from other papers, if you focus on the Introduction and the Related Work sections, you should find 70% of the hypotheses from other papers. The results can be different for the specific paper you are analyzing. The table is only a summary of previous results.

## 5) Data Entry Forms

An excel file contains worksheets for recording:

- Context descriptions
- Results

### 5.1) Context Descriptions

Fill out one form for each paper (or each study recorded in the paper, if there are multiple studies). The attributes of the context description form are:

- **Paper Title:**

- The title of the paper from which you are extracting the information.

- **Topic:**

- We use the IEEE keywords from the Computer.org website to denote topic categories. The main topics are the following, but this list can be extended using the extended list on (<http://www.computer.org/mc/keywords/software.htm>):

- Software Engineering – General
- Requirements/Specifications
- Design Tools and Techniques
- Coding Tools and Techniques
- Software/Program Verification
- Testing and Debugging
- Programming Environments/Construction Tools
- Distribution, Maintenance, and Enhancement
- Metrics/Measurement
- Management
- Design
- Software Architectures
- Interoperability
- Reusable Software
- Human Factors in Software Design
- Software and System Safety
- Configuration Management
- Software Construction
- Software Engineering Process
- Software Quality/SQA

- **Type of the Study**

- **Experiment** - A detailed and formal investigation that is performed under controlled conditions with the objective to manipulate one or more variables (called independent variables or the variables under study) and control all other variables at fixed levels. The purpose of a controlled experiment is to make observations whose causes are unambiguous. This is achieved by isolating the effects of each factor (the dependent variables) from the effects of other factors to make significant claims of cause and effect.
- **Case Study** - A detailed investigation of a single “case” or a number of related “cases”. Such an investigation is performed under normal conditions, e.g., a representative project in some organization. In a case study, the variables are not controlled for but identified as they exist.

- **Survey** - A broad investigation where information is collected in a standardized form from a group of people or projects. The primary means of gathering qualitative or quantitative data are interviews or questionnaires.
- **Goals:**
  - States de goals for the study described on the paper using the GQM goal template using the form:
    - **Analyze object of study**
      - Object of study (the attributes of the entities that are of interest, the purpose of the study i.e., the process, product, model, metric, ...)
    - **for the purpose of X**
      - X (i.e., whether the study is aimed at characterizing, understanding, evaluating, predicting, or improving)
    - **with respect to M**
      - M (example: effectiveness, number of defects.
    - **from the point of view of P**
      - P (for whom the study should be of value, i.e., a researcher, project manager, corporation, ...).
  - Example of goal:

<b>Goals</b>	<i>Analyze</i>	code reading, functional testing, sructural testing
	<i>for the purpose of</i>	evaluating
	<i>with respect to</i>	# failures observed & time per fault
	<i>from the point of view of</i>	researcher

- **Variables**
  - A variable is a concept or construct that can vary or have more than one value. The researcher might then be interested in knowing how certain variables are related to each other. For example, which variables predict “effectiveness” of a testing technique? Or, he might be interested in understanding the relationship between the number of defects and the size of the programs.
  - There are two basic kinds of variables.
    - The independent variable is defined by these authors as the "variable that the experimenter manipulates." While this is true in experiments, not all studies are experiments. Often, researchers don't manipulate anything in a study. Instead, they merely collect data and observe how variables are related to each other. The independent variable is what the researcher is studying with respect to how it is related to or influences other variables (the dependent variables). If the independent variable is related to or influences the dependent variable, it can be used to predict the dependent variable. It is therefore sometimes called the predictor variable, or the explanatory

variable. The independent variable may be manipulated or it may just be measured.

- In contrast, the dependent variable is what the researcher is studying, with respect to how it is related to or influenced by the independent variable or how it can be explained or predicted by the independent variable. It is sometimes called the response variable or the criterion variable. It is never manipulated as a part of the study.
- A useful hint for determining which variable is which in a study is to ask whether you are trying to either influence or predict one variable from some other variable or variables. If so, the variable you are trying to predict is probably the dependent variable. The variable that you are using to make the predictions or to determine if it influences (rather than is influenced by) some other variable in the study is typically the independent variable.
- Describe as many as possible of the following characteristics for each dependent and independent variable in the study:
  - *Name*: How the variable is referred to in the paper.
  - *Type*: Type of the variable: independent, dependent, both (Dependent and Independent), unclear;
  - *Possible Values*: The possible values for the variable, if controlled.
  - *Data Collection Details*: Details of the method used to measure the variable, including for example what instrumentation and tool support were used.

#### • **Subjects**

- Describe as many as possible of the following characteristics for the executors of the study:
- If the experiment is totally automatic, a tool must be the executor of the task and probable will be described at the section Instrumentation. In this case this field should be left blank.
- If the experiment is not totally automatic a subject must be the executor of the task:
  - *Category*: A generalized description of the experience level of subjects. Possible values here can be:
    - Undergraduate Students;
    - Graduate Students;
    - Students: This is an “unknown” type of students.
    - Professionals;
    - Scientists;
    - Other. Specify.
    - Unknown: Not described in the paper.
  - *Number*: The number of subjects that participated in the experiment.

#### • **Instrumentation**

- Automatic Measurements Tools, auxiliary tools that generates data for the experiment. It is used in the task to generate data.

- *Tool Name*: name of the program
  - *Description of Functions*: description of the functions performed on the experiment.
- **Task**
  - *Category*: Categorize the tasks given to the “executors” according to the **tasks** applied and the **work products** they were applied to (e.g. *create a design document*).
    - Possible values of tasks:
      - Plan
      - Create
      - Modify
      - Analyze
    - Possible work products:
      - Requirements
      - Architecture/design
      - Code
      - Change Reports
      - Error Reports
      - Etc.
  - *Duration*: Duration of time that executors used to perform the task(s).
  - *Work Mode*: Select whether subjects performed the task(s) as:
    - Team
    - Individual
- **Work Products**
  - Description of the working products used in the task. They may be, for example, specification or code documents. Usually each experiment uses more than one instrument.
  - *Name*: A name for the product, usually the name on which it is referenced in the text.
  - *Type*: Possible types of work products:
    - Requirements
    - Architecture/design
    - Code
    - Change Reports
    - Error Reports
    - Other, specify.
  - *Application Origin*: The origin of application where the tasks are performed on. The possible values are:
    - Constructed: Applications constructed for the purpose of the experiment;
    - Commercial: A commercial application;
    - Student Project: An application constructed for a class assignment;
    - Open Source: An open source application;

- Other. Specify.
  - Unclear
- *Application Domain:*
  - Text Processor
  - Flight Simulation
  - Etc...
- *Size:*
  - Size of the application, using the metric specified by the author. Ex: 129 lines of code, of 2000 executables lines of code, etc
- *Representation Paradigm*
  - The representation paradigm of the work product. Example: Object Oriented, Imperative, Structured.
- *Language*
  - Language used to write the work product. Example, English, Fortran, Pascal, C++, Java.
- **Replication:** Indicate whether this study is a replication of another one. (Choose “yes” or “no.”) Include a reference to the original experiment if this is a replication. And the differences between the replications.
- **Other:** Note any other information that is important for understanding the model, metric, techniques, or the empirical study itself (e.g., missing definitions, environmental characteristics, or information about process conformance).

## 5.2) Results

Fill out one form for each result you identify. The attributes of the results form are:

- **Plain Text:**
  - Try to write the result using the words from the paper so that traceability is assured. When identifying results, recall that: The results should be stated in such a way that data can be collected that either supports or refutes it. A good result states as clearly and concisely as possible the expected relationship (or difference) between dependent and independent variables and defines those variables in operational, measurable terms.
- **Origin**
  - **Section name, Figure or Table reference on which the result was gathered**
- **Type:**
  - Tested Result
  - Result from another paper (A result that was tested in another paper and reported in this one.)
- **Level of Support:** The support should be described as one of the following levels:
  - Significantly positive: The results are statistically significant, that is, with a high degree of certainty are not resultant of pure chance.

- Positive: The data in the paper support the result, but no significant statistical results can back this up.
- Null: The data in the paper neither support nor contradict the result.
- Belief: The statement is formulated based on assumption or belief but has not been tested.
- **Observations:**
  - This is a free-text field for you to keep track of any additional information that is important for correctly understanding or interpreting the results.

# Apêndice B

## Entrevista com Especialistas

### B.1 - Entrevista com Especialista I

*1. Quais foram os critérios de seleção dos artigos? Especifique critérios para a inclusão e a exclusão dos artigos.*

No começo, a pesquisa estava relacionada à sua tese de doutorado; portanto, foi ela quem fez a busca por todos os artigos. Primeiro ela tentou buscar todos os artigos de teste que continham algum experimento. Primeiramente a busca foi na web, usando palavras-chaves como experimentos de teste, estudos empíricos em teste. A partir desta busca, um conjunto de artigos foram encontrados. Em seguida ela fez o que chamou de “closure”, para cada artigo, ela analisou as referências dos artigos para completar o conjunto de artigos. Com esta busca, ela acredita que conseguiu uma lista mais abrangente de artigos. Ela acredita que esta prática de olhar nas referências é importante porque às vezes pensamos que a busca na Internet, catálogos e assim por diante está completa, mas sempre alguns artigos adicionais são encontrados na revisão das referências. A qualidade da publicação foi um dos critérios para a inclusão de artigos; um dos critérios era que a publicação deveria ser de uma conferência ou periódico conhecido na área. Alguns artigos não foram incluídos porque ela não conseguiu encontrar o artigo. Na seleção dos artigos elas removeram um artigo, pois este não continha detalhes suficientes sobre o experimento relatado no artigo. O artigo não dizia, por exemplo, quais técnicas eles investigavam no estudo e omitiram alguns detalhes do processo experimental. Havia uns 4 ou 5 artigos que eram estudos isolados e como elas tentaram criar grupos, elas não conseguiram incluir estes 4 ou 5 artigos nos grupos e, portanto elas acharam que não poderiam extrair conclusões deles por agregação.

*2. Como foi a divisão dos artigos? Cada uma de vocês revisou o artigo individualmente ou em pares? Existia uma ordem pré-definida para a leitura dos artigos?*

Ana Moreno e Sira Vegas extraíram as informações dos artigos, em pares. Sira Vegas fez primeiramente uma revisão dos artigos para a sua tese de doutorado; o objetivo não era exatamente o mesmo que o do artigo. Assim, algumas informações elas poderiam ter reusado, mas elas acreditam que muito pouco destas informações poderiam ser reaproveitadas; portanto, elas decidiram que iriam reler todos os artigos. Para a releitura dos artigos, sempre em pares elas sentavam e liam o artigo ao mesmo tempo e trocavam informações sobre os artigos. Algumas vezes uma delas não havia compreendido a informação. Às vezes ambas haviam compreendido diferentemente as informações contidas nos artigos e elas tiveram que ler novamente o artigo para tentar chegar a um consenso; às vezes elas precisavam de uma terceira pessoa para chegarem a um consenso. Ela acredita que esta é uma boa prática e acredita ser importante ter mais de uma pessoa lendo o artigo. Ela acredita que depende de quão bem o artigo foi escrito, alguns artigos escritos há 20 ou 30 anos naturalmente não incluíram toda a informação necessária para detalhar a conteúdo um experimento.

### ***3. Havia um processo específico para ler cada artigo?***

Para ela alguns artigos eram mais fáceis de ler. Como a Sira tinha lido os artigos para o doutorado, ela tinha como classificar os artigos. Assim elas decidiram começar com os artigos que eram mais fáceis de analisar. Os mais complicados foram deixados para o final. Em geral os artigos mais detalhados eram mais fáceis. Algumas vezes, só de analisar a estrutura do artigo elas conseguiam dizer se este continha as informações que elas necessitavam; além disto, a Sira já sabia quais eram as partes importantes dos artigos então ela muitas vezes direcionava a leitura para as seções mais importantes dos artigos.

### ***4. Havia uma informação específica que vocês procuravam em cada artigo?***

Sobre o contexto do experimento elas procuravam primeiramente as variáveis dependentes e as técnicas, estas eram as informações fáceis de encontrar. Mas, extrair as conclusões foi muito mais difícil. Alguns artigos descreviam claramente as conclusões, outros não. Algumas

vezes os artigos não descreviam os resultados ou conclusões que poderiam ser estabelecidos sobre o estudo descrito. Nestes casos, elas tinham que ler os gráficos, figuras, etc. e abstrair as conclusões. Para ela, em alguns artigos, mesmo atualmente, os leitores ainda necessitam fazer esta abstração.

**5. *Você tem a idéia do tempo gasto em cada artigo?***

Dependeu muito do artigo. Alguns artigos eram fáceis de ler, e às vezes faziam a análise entre meia e uma hora. Os artigos mais complicados podiam levar até 3 horas. Esta estimativa é para a rodada de análise em pares.

**6. *Que torna um artigo difícil ou fácil de ler?***

O mais difícil segundo ela é interpretar os resultados descritos nos artigos. A outra especialista ajudou muito no refinamento das conclusões. Ela ainda acredita que se outra pessoa fizer a mesma análise nos artigos pode chegar a conclusões diferentes das delas. E não dá pra saber quem estaria certo ou errado. Talvez somente quem escreveu o artigo pode dizer ao certo que ele queria dizer no texto. Ela acredita que mesmo não sendo muito precisas, elas acreditam que deveriam fazer uma tentativa.

**7. *Quanto sua experiência com teste ajudou na análise?***

Ela acredita que para fazer a agregação, a pessoa necessita conhecer o tópico. Quando as conclusões estão lá no artigo é fácil ler e extrair a informação do artigo. Os artigos bem escritos são mais fáceis e o leitor não necessita interpretar os resultados. O detalhamento do experimento é um dos fatores que o torna mais fácil de ler. Sem a experiência em teste ela não acredita que seria possível fazer a extração. Mas que não acha que necessita-se de um especialista no tópico para a extração.

**8. *Quanto sua experiência com experimentação ajudou na análise?***

Para ela a experiência em experimentação foi ainda mais importante do que a experiência em teste. Naturalmente isto pode estar relacionado a como os artigos foram escritos, mas ela diria que, para ler estes artigos, é preciso saber o significado de, por exemplo, variáveis dependentes, independentes e de variáveis do contexto. Ela disse que as variáveis dependentes e independentes são fáceis de reconhecer, e definitivamente as variáveis dependentes são mais fáceis. Mas para os outros aspectos do contexto, é necessário o conhecimento de experimentação para extrair a informação.

***9. Para você, qual é a definição de “resultado”? Um resultado válido tem sempre que ser baseado em dados?***

Os resultados são um pedaço de conhecimento (piece of knowledge). Elas não consideraram opiniões somente porque ela acredita que esta informação não é útil para profissionais.

***10. Quando você estava analisando um artigo, você tentou extrair sempre todos os resultados que ele poderia ter, ou você teve um objetivo específico nos tipos de resultados que você quis coletar?***

Esta resposta já está implícita em outras respostas. Elas tentaram coletar as conclusões mais importantes do texto.

***11. Como você tratou resultados de outros artigos? Você os considerou em sua análise?***

Para evitar equívocos, elas evitaram coletar resultados de outros artigos. Quando estes resultados eram encontrados elas tentavam incluir o artigo na análise, quando possível, e analisar o artigo em questão. Algumas vezes estes “resultados de outros artigos” eram usado como uma certificação de que elas entenderam corretamente os resultados descritos no artigo.

***12. Você leu o artigo inteiro ou apenas algumas seções específicas para procurar pela informação que você necessitava?***

A Sira leu os artigos para a tese de doutorado. Quando ela analisou os artigos com a Ana Moreno, elas fizeram uma leitura diagonal, e analisava com mais detalhes as seções que pareciam ser importantes em cada artigo. Mas quando o artigo era um pouco mais complicado elas precisaram ler o artigo mais de uma vez.

***13. Para avaliar as limitações. Vocês usaram uma abordagem específica para escolher o que seria analisado? Por exemplo: tipo da análise, da categoria de assuntos etc. Se sim, você usou esta abordagem para todos os artigos?***

Não existia uma lista específica, basicamente elas procuravam saber se os autores usaram programas reais ou não; no princípio elas acharam que daria para fazer a lista à medida que fossem lendo os artigos. Assim, elas não estabeleceram uma lista de características a priori, mas depois, a lista foi sendo criada incrementalmente e tiveram que re-visitar os artigos para que tivessem as mesmas informações sobre todos os artigos. Para ela é mais uma questão de tempo para extrair algumas informações de contexto do que de conhecimento requerido.

***14. Sabemos que vocês extraíram as variáveis dependentes dos estudos e as técnicas de testes usadas em cada estudo. Você acredita ser importante avaliar as variáveis independentes dos estudos?***

Não focaram em todas as variáveis independentes, somente nas que eram importantes para elas. Os estudos eram tão heterogêneos, que era difícil abstrair o que existia em comum entre eles. Basicamente elas extraíram as técnicas e às vezes a experiência dos participantes.

*15. Sobre os resultados práticos, vocês tiveram um processo para validá-los? Como contagem de votos etc.?*

Não.

## **B.2 - Entrevista com Especialista II**

*1. Quais foram os critérios de seleção dos artigos? Especifique critérios para a inclusão e a exclusão dos artigos.*

A especialista falou das duas experiências que ela já teve com agregação de estudos: “Reviewing 25 years of testing technique experiments” [100] e “Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review” [4]. Neste segundo, os autores tentaram ser um pouco mais formal do que elas no artigo de teste, eles também estão tentando formalizar o processo.

Artigo de Teste: Foi um processo mais informal, a busca veio da tese da Sira. Ela tentou ser o mais abrangente possível, sem focar em alguma conferência específica. Ela acredita que a melhor maneira é fazer uma busca na web, mas que nem todos os artigos contribuem para o estudo e que alguns deles são removidos por causa dos problemas da qualidade. Ela disse que é muito difícil dizer que o artigo foi removido por causa da qualidade, e este é um problema persistente. Elas tentaram desenvolver critérios de qualidade ou quantidade de informações. A exclusão segundo ela é possível muitas vezes somente após ter lido o artigo.

Artigo sobre Requisitos: Eles realizaram a busca em todas as bases de dados, além disso eles fizeram a mesma análise de referências que fizeram no outro artigo e analisaram a quantidade de referências incluídas com esta análise. Ela mencionou que esta foi uma tarefa muito difícil e somente poucos artigos foram incluídas. Ela acredita que não vale a pena fazer esta análise de referências, para ela, realizar a busca em bases de dados é o melhor em termos de esforço e ganho. Se ela fosse recomendar algo, diria para não fazer exame das referências como uma fonte de artigos. Sobre os artigos “candidatos” e os artigos escolhidos, ela acredita

que deve-se ler rapidamente o texto para ver se ele deve ser escolhido ou não. Depois deve-se ler realmente e avaliar a qualidade do artigo.

**2. *Como foi a divisão dos artigos? Cada uma de vocês revisou o artigo individualmente ou em pares? Existia uma ordem pré-definida para a leitura dos artigos?***

Artigo de Teste: Elas classificaram os artigos como famílias, agrupando-os de acordo com as técnicas de teste. O processo teve um único passo, ler, obter a informação e classificação. O processo foi feito em pares, mas não em todo o tempo. A especialista I tinha lido a priori, e depois a Ana Moreno (Especialista III) leu os mesmos artigos e elas fizeram reuniões para verificar se as duas tiveram a mesma compreensão do artigo. Ela foi a “conciliadora” do processo. Ela disse que é muito difícil saber se as extrações foram confiáveis. A linguagem natural nem sempre é clara. Ela acredita que é necessário que mais de uma pessoa realize o trabalho de extração.

Artigo sobre Requisitos: Neste artigo elas tentaram formalizar mais a extração por causa dos problemas que tiveram antes. No artigo de teste eles perderam a rastreabilidade dos resultados em relação ao texto no qual estes foram extraídos.

**3. *Havia um processo específico para ler cada artigo?***

Como ela não participou desta parte do processo, ela disse que não tem como responder a esta questão, mas que tem certeza de que eles não leram o artigo como tradicionalmente o fazem.

**4. *Havia uma informação específica que vocês procuravam em cada artigo?***

Sim, no caso do artigo de requisitos eles definiram uma tabela antes de ler o artigo. Na opinião dela é necessário ter-se um modelo para buscar por esta informação.

**5. *Você tem a idéia do tempo gasto em cada artigo?***

Artigo de Teste: Ela não tem idéia.

Artigo de requisitos: Ela acredita que é difícil de mensurar no total o tempo gasto em cada artigo, porque existe sempre a volta à fonte das informações. Não é somente extrair as informações, pois as tabelas de extração não estão completas e então é sempre necessário sempre voltar ao artigo para que algumas dúvidas sejam sanadas.

**6. *Que torna um artigo difícil ou fácil de ler?***

Existem artigos bons ou ruins, completos ou incompletos. Os piores casos são de artigos incompletos. Porque você permanece lendo o artigo para ver se a informação está lá receosa de que ela está lá, mas você não consegue encontrá-la. A estrutura do papel ajuda muito e os artigos curtos são mais difíceis de ler.

**7. *Quanto sua experiência com teste ajudou na análise?***

As respostas às questões 7 e 8 estão na resposta à questão 8.

**8. *Quanto sua experiência com experimentação ajudou na análise?***

A principal questão é que o melhor caso é quando o leitor é um perito na experimentação e perito no assunto que está revisando. Algumas das discussões que elas tiveram eram porque a Ana Moreno não era perita em testes. Na extração das informações é necessário estar confortável com o assunto. Para Oscar no artigo de Requisitos era fácil extrair as informações porque ele se sente mais confortável com o assunto tratado nos artigos apesar de Oscar ter tido problema com a experimentação no início. Daí ele teve que parar a leitura dos artigos para estudar sobre experimentação. O que a faz crer que ao ler e ao extrair a informação do artigo, o conhecimento do domínio não é tão importante quanto a experiência

com experimentação. Mas, para a agregação, a experiência do domínio é mais importante. Talvez, tão importante quanto a experiência com a experimentação.

**9. *Para você, qual é a definição de “resultado”? Um resultado válido tem sempre que ser baseado em dados?***

O resultado é a saída de um experimento. Não são os dados. Poderia ser o resultado da análise naqueles dados. Para ela, um dos problemas nos artigos é que a maioria dos autores não interpreta os seus resultados, quase tudo que é descrito nos artigos é muito orientado a dados. Para ela, para se chegar a uma evidência existe um grande gap. Ela acredita que existe sempre a necessidade do experimentador em fazer algum tipo de interpretação.

Para esta especialista, a evidência é algo mais. Para ela, um resultado é o que um experimentador encontra, não precisa de muita interpretação. O Victor Basili e também a especialista acreditam que a evidência começa a existir somente quando alguma síntese foi feita nos resultados. Um pedaço de conhecimento, que é quando se realmente amadurece sobre um conhecimento.

Nas análises que elas executaram, elas excluíram todos aqueles artigos sem análise de dados formal.

**10. *Quando você estava analisando um artigo, você tentou extrair sempre todos os resultados que ele poderia ter, ou você teve um objetivo específico nos tipos de resultados que você quis coletar?***

Não. Esta é a razão que o modelo é importante para recolher a informação. Mas não é necessário extrair tudo somente o que é importante para alguma questão que se está tentando responder. Um artigo pode contribuir para muitas revisões sistemáticas. Ela acredita que extrair tudo na primeira etapa e esquece-se então que o artigo existe não é bom. Ela acredita que podemos reusar o que já foi extraído, mas eles ainda não tiveram este caso e que estamos

incluindo uma etapa nova antes do início de revisão sistemática concreta. Ela fala de um outro cenário no qual uma base de informações pode ser usada. Quando se tem muitos artigos extraídos em uma base de dados, então você pode criar uma visualização e então ver onde não há nenhum estudo, ou formular perguntas baseadas no que está na base já sobre alguma área específica como por exemplo: Que variáveis do contexto estão cobertas?

Ela levantou a questão de se realizar todo o processo sem auxílio de ferramentas, e que manter a rastreabilidade de todo processo é muito difícil apesar de ser uma tarefa muito importante, para todo o processo e também para as questões de reuso da base.

***11. Como você tratou resultados de outros artigos? Você os considerou em sua análise?***

Elas não trataram resultados de outros artigos e também as opiniões não foram consideradas. Quando existiam casos de resultados de outros artigos, elas buscavam incluir estes artigos na análise.

***12. Você leu o artigo inteiro ou apenas algumas seções específicas para procurar pela informação que você necessitava?***

Lemos todo o artigo.

***13. Para avaliar as limitações. Vocês usaram uma abordagem específica para escolher o que seria analisado? Por exemplo: tipo da análise, da categoria de assuntos etc. Se sim, você usou esta abordagem para todos os artigos?***

Sim. Tipo de análise, e elas excluíram os artigos que não formalizaram os resultados baseados numa análise estatística.

No artigo de requisitos, eles estão fazendo o processo em duas etapas, os artigos com qualidade alta e os artigos com qualidade baixa. Os artigos com qualidade baixa ratificam sobre as evidências da alta qualidade. Ela diz que às vezes há dois resultados discordantes,

mas um está em um experimento muito bom, o outro está em um experimento muito ruim, e então, o que deve ser feito? Ela diz não ter uma resposta concreta a esta questão, mas que às vezes descarta a evidência com qualidade ruim.

***14. Sabemos que vocês extraíram as variáveis dependentes dos estudos e as técnicas de testes usadas em cada estudo. Você acredita ser importante avaliar as variáveis independentes dos estudos?***

Elas usaram as variáveis dependentes dos estudos para agrupá-los. Mas, no agrupamento dos resultados eles usaram todos os fatores. Ela comenta sobre a questão da inconsistência nos resultados. E diz que a questão dos termos e definição de termos é um grande problema. Eles tentam abstrair termos e criar categorias de termos que são sinônimos.

***15. Sobre os resultados práticos, vocês tiveram um processo para validá-los? Como contagem de votos etc.?***

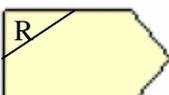
O processo era ainda imaturo no primeiro artigo. Agora eles usam algum tipo de contagem do voto. Fizeram alguns resumos e listas dos resultados que encontraram para criar algumas evidências. Para ela, esta é a parte mais imatura do trabalho que eles têm desenvolvido.

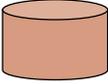
# Apêndice C

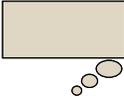
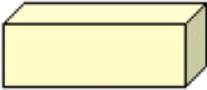
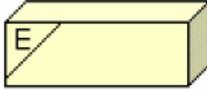
## Notação utilizada na Modelagem da SecESE

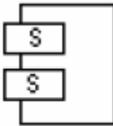
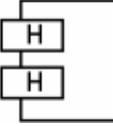
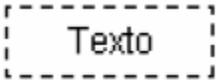
### C.1 – Notação

Este anexo apresenta a notação utilizada na modelagem do processo de construção interativa de árvores de classificação proposto neste trabalho. [71]

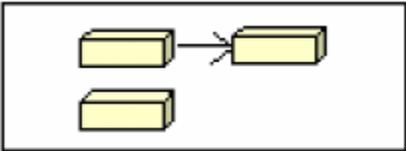
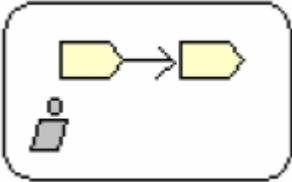
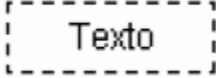
Entidade	Forma de Representação
Estado Inicial	
Estado Final	
Ator	
Atividade Atômica	
Atividade Composta	
Atividade Implícita	
Atividade de Registro	

Atividade Externa	
Repositório (Meio Magnético)	

<b>Entidade</b>	<b>Forma de Representação</b>
Conhecimento Explícito	
Conhecimento Tácito	
Comunicação	
Processo	
Processo Externo	
Evento	
Arquivo (Local Físico)	
Documento	

Componente de Software	
Componente de Hardware	
Peça	
Matéria Prima	
Bem	
Nota Explicativa	

Operações Lógicas	
AND	
OR	
XOR	
Fluxo de Controle	

Fluxo de Entrada / Saída	
Associação não Direcionada	
Associação de Nota Explicativa	
Grupo de Processo	
Área de Ator	
Nota Explicativa	

# **Apêndice D**

## **Dicionário de Dados da Base de Resultados**

Tabela: AnalysisContext

**Colunas**

Nome	Tipo	Tamanho
AnalysisID	Texto	50
Analyst ID	Texto	250
GOAL	Memorando	-
Initial Hypothesis	Memorando	-
AnalysisMethodology	Memorando	-
AnalysisofCompleteness	Memorando	-
SelectionCriteria	Memorando	-
TypesofStudies to Be Included	Texto	50
TypesofParticipants to Be Included	Texto	200
Forms and Type of Publication to Be Included	Memorando	-
PotencialBiases	Texto	250
DateofTheAnalysis	Texto	50

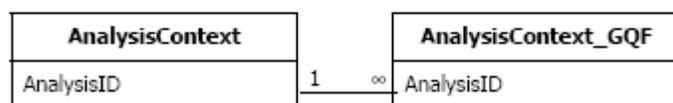
**Relacionamentos**

**AnalysisContextAnalysisXQualityCriteriaXPaper**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**AnalysisContextAnalysisContext\_GQF**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**AnalysisContextAnalysisXFonts**

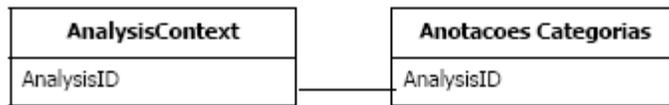


Attributes: Imposto  
 RelationshipType: Um-para-muitos

Tabela: AnalysisContext

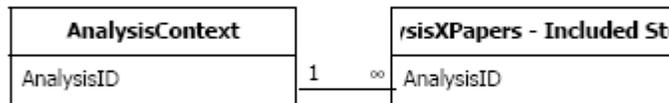
!

#### AnalysisContextAnotacoes Categorias



Attributes: Não imposto  
RelationshipType: Um-para-muitos

#### AnalysisContextAnalysisXPapers - Included Studies



Attributes: Imposto  
RelationshipType: Um-para-muitos

#### AnalysisContextConclusionsXPapers



Attributes: Imposto  
RelationshipType: Um-para-muitos

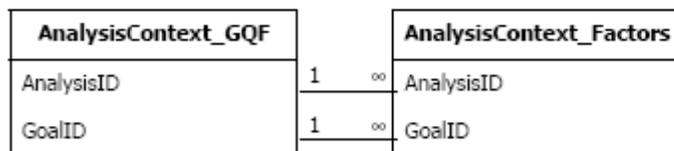
Tabela: AnalysisContext\_Factors

**Colunas**

Nome	Tipo	Tamanho
Number	Inteiro longo	4
AnalysisID	Texto	50
GoalID	Byte	1
factor_name	Texto	50
factor_def	Inteiro longo	4
factor_type	Texto	50

**Relacionamentos**

**AnalysisContext\_GQFAnalysisContext\_Factors**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

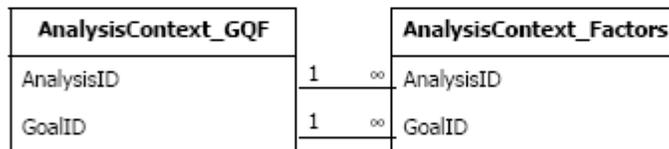
Tabela: AnalysisContext GOF

**Colunas**

Nome	Tipo	Tamanho
AnalysisID	Texto	50
GoalID	Byte	1
Object of Study	Texto	50
purpose	Texto	50
ScopeConstraint_Type_of_Study	Texto	50
ScopeConstraint_Goal_Object_of_Study	Texto	255
ScopeConstraint_Subjects_category	Texto	50
ScopeConstraint_subjects_numbers	Inteiro longo	4
ScopeConstraint_subjects_work_mode	Texto	50
ScopeConstraint_task_category	Texto	50
ScopeConstraint_task_category_workProducts	Texto	50
ScopeConstraint_task_duration	Texto	50
ScopeConstraint_instrument_type	Texto	50
ScopeConstraint_instrument_origin	Texto	50
PointofView	Texto	50

**Relacionamentos**

**AnalysisContext\_GQFAnalysisContext\_Factors**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**AnalysisContextAnalysisContext\_GQF**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

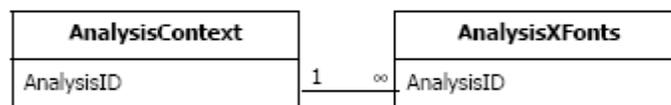
Tabela: AnalysisXFonts

### Colunas

Nome	Tipo	Tamanho
AnalysisID	Texto	50
FontID	Inteiro longo	4
SearchCriteria	Memorando	-

### Relacionamentos

#### AnalysisContextAnalysisXFonts



Attributes: Imposto  
RelationshipType: Um-para-muitos

#### FontesAnalysisXFonts



Attributes: Imposto  
RelationshipType: Um-para-muitos

Tabela: AnalysisXPapers - Included Studies

**Colunas**

Nome	Tipo	Tamanho
PaperID	Inteiro longo	4
AnalysisID	Texto	50
IDRelation	Inteiro longo	4
Included?	Sim/Não	1
FontID	Inteiro longo	4

**Relacionamentos**

**AnalysisContextAnalysisXPapers - Included Studies**



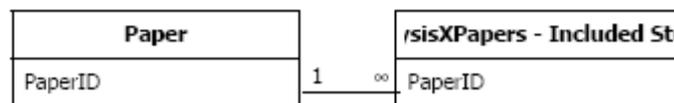
Attributes: Imposto  
 RelationshipType: Um-para-muitos

**FontesAnalysisXPapers - Included Studies**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperAnalysisXPapers - Included Studies**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

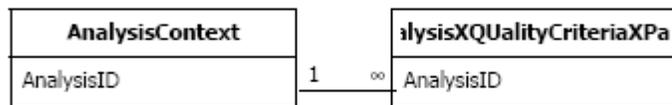
Tabela: AnalysisXQualityCriteriaXPaper

**Colunas**

Nome	Tipo	Tamanho
AnalysisID	Texto	50
CriteriaID	Inteiro longo	4
PaperID	Inteiro longo	4
Value	Inteiro longo	4

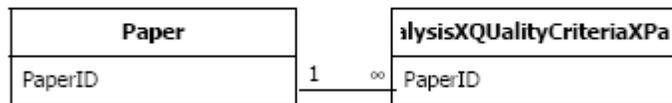
**Relacionamentos**

**AnalysisContextAnalysisXQualityCriteriaXPaper**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperAnalysisXQualityCriteriaXPaper**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**QualityCriteriaAnalysisXQualityCriteriaXPaper**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

Tabela: Anotacoes Cateoarias

**Colunas**

Nome	Tipo	Tamanho
NoteID	Inteiro longo	4
AnalysisID	Texto	50
NoteType	Texto	100
Title	Texto	255
Description	Memorando	-
Decision/Causa do Problema/Fatos que suportam a ideia/Origem	Memorando	-
Justificativa da Decisão/ConseqüênciasFato/OutrasSituçõesAplica	Memorando	-
Solução	Memorando	-
Resultados	Memorando	-

**Relacionamentos**

**AnalysisContextAnotacoes Categorias**



Attributes: Não imposto  
 RelationshipType: Um-para-muitos

Tabela: Authors

**Colunas**

Nome	Tipo	Tamanho
AuthorID	Inteiro longo	4
AuthorName	Texto	255

**Relacionamentos**

**AuthorsPaperXAuthors**



Attributes:  
RelationshipType:

Imposto  
Um-para-muitos

Tabela: Conclusions

**Colunas**

Nome	Tipo	Tamanho
ConclusionID	Inteiro longo	4
ConclusionDescription	Memorando	-
Annex	Objeto OLE	-
DefinitionReferenceID	Inteiro longo	4
ConclusionsUncertainties	Memorando	-

**Relacionamentos**

**AnalysisDefinitionsReferenceConclusions**



Attributes: Não imposto  
 RelationshipType: Um-para-muitos

**ConclusionsConclusionsXPapers**



Attributes: Não imposto  
 RelationshipType: Um-para-muitos

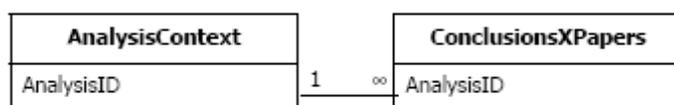
Tabela: ConclusionsXPapers

**Colunas**

Nome	Tipo	Tamanho
ConclusionID	Inteiro longo	4
PaperID	Inteiro longo	4
AnalysisID	Texto	50
Agree with Conclusion?	Texto	50
Attenuations	Memorando	-

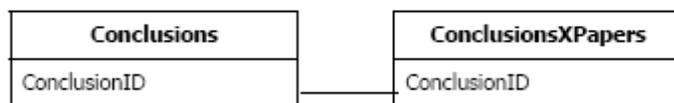
**Relacionamentos**

**AnalysisContextConclusionsXPapers**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**ConclusionsConclusionsXPapers**



Attributes: Não imposto  
 RelationshipType: Um-para-muitos

**PaperConclusionsXPapers**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

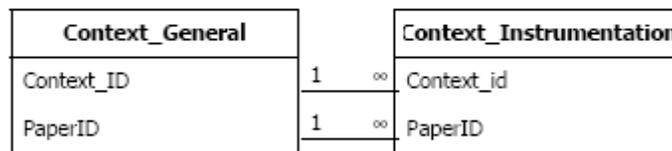
Tabela: Context General

**Colunas**

Nome	Tipo	Tamanho
Context_ID	Inteiro longo	4
PaperID	Inteiro longo	4
Type_of_Study	Texto	50
Topic	Inteiro longo	4
Replication_reference	Texto	255
Goal_description	Memorando	-
Goal_Object_of_Study	Texto	255
Subjects_category	Texto	50
subjects_numbers	Inteiro longo	4
subjects_work_mode	Texto	50
task_category	Texto	50
task_category_workProducts	Texto	50
task_duration	Texto	50

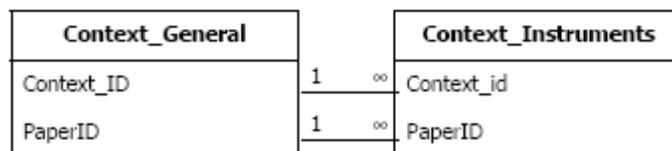
**Relacionamentos**

**Context\_GeneralContext\_Instrumentation**



Attributes: Imposto, Propagar exclusões  
 RelationshipType: Um-para-muitos

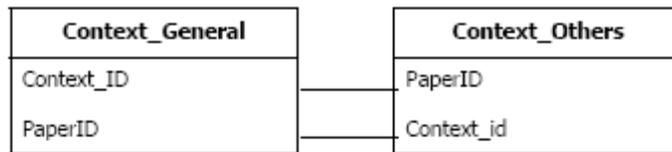
**Context\_GeneralContext\_Instruments**



Attributes: Imposto, Propagar exclusões  
 RelationshipType: Um-para-muitos

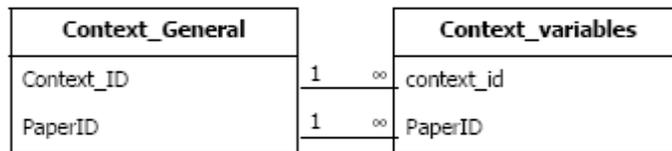
Tabela: Context General

**Context\_GeneralContext\_Others**



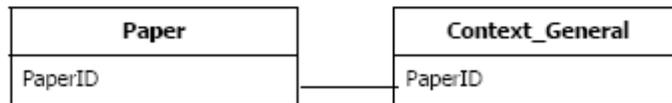
Attributes: Não imposto  
RelationshipType: Um-para-muitos

**Context\_GeneralContext\_variables**



Attributes: Imposto, Propagar exclusões  
RelationshipType: Um-para-muitos

**PaperContext\_General**



Attributes: Não imposto  
RelationshipType: Um-para-muitos

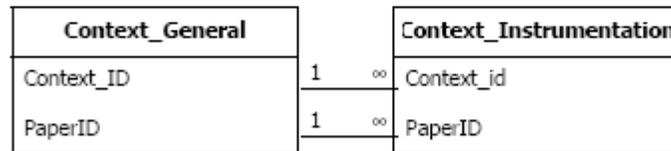
Tabela: Context Instrumentation

**Colunas**

Nome	Tipo	Tamanho
Number	Inteiro longo	4
PaperID	Inteiro longo	4
Context_id	Inteiro longo	4
Instrumentation_name	Texto	50
instrumentation_description	Memorando	-

**Relacionamentos**

**Context\_GeneralContext\_Instrumentation**



Attributes: Imposto, Propagar exclusões  
 RelationshipType: Um-para-muitos

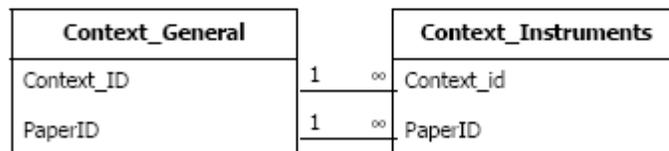
Tabela: Context Instruments

**Colunas**

Nome	Tipo	Tamanho
Number	Inteiro longo	4
PaperID	Inteiro longo	4
Context_id	Inteiro longo	4
Instrument_name	Texto	50
instrument_type	Texto	50
instrument_origin	Texto	50
instrument_application_domain	Memorando	-
instrument_size	Texto	50
instrument_representation_paradigm	Texto	50
instrument_language	Texto	50

**Relacionamentos**

**Context\_GeneralContext\_Instruments**



Attributes: Imposto, Propagar exclusões  
 RelationshipType: Um-para-muitos

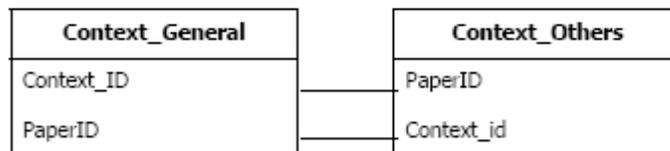
Tabela: Context\_Others

**Colunas**

<u>Nome</u>	<u>Tipo</u>	<u>Tamanho</u>
Number	Inteiro longo	4
PaperID	Inteiro longo	4
Context_id	Inteiro longo	4
others_category	Texto	50
others_values	Memorando	-

**Relacionamentos**

**Context\_GeneralContext\_Others**



Attributes: Não imposto  
RelationshipType: Um-para-muitos

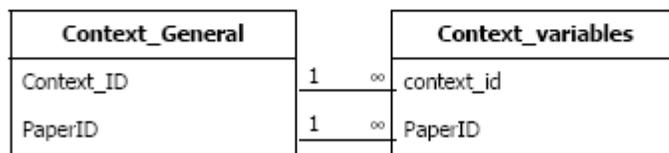
Tabela: Context variables

**Colunas**

Nome	Tipo	Tamanho
Number	Inteiro longo	4
PaperID	Inteiro longo	4
context_id	Inteiro longo	4
variable_type	Texto	50
variable_name	Texto	50
variable_def	Inteiro longo	4
variable_possible_values	Memorando	-
variable_data_collection_details	Memorando	-

**Relacionamentos**

**Context\_GeneralContext\_variables**



Attributes: Imposto, Propagar exclusões  
 RelationshipType: Um-para-muitos

Tabela: Fontes

### Colunas

Nome	Tipo	Tamanho
FontID	Inteiro longo	4
Font Name	Texto	200
FontURL	Texto	255

### Relacionamentos

#### FontesAnalysisXFonts



Attributes: Imposto  
RelationshipType: Um-para-muitos

#### FontesAnalysisXPapers - Included Studies



Attributes: Imposto  
RelationshipType: Um-para-muitos

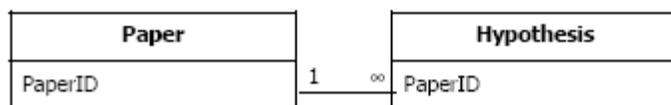
Tabela: Hypothesis

**Colunas**

Nome	Tipo	Tamanho
PaperID	Inteiro longo	4
Id	Inteiro longo	4
Number	Inteiro longo	4
Origin/Section/Figures/Tables	Texto	50
Plain_Text	Memorando	-
Weight-Hypothesis/Belief/Opinion	Inteiro longo	4
Hypothesis/Belief/Opinion	Texto	100
Control_I	Texto	100
Magnitude_of_Difference_I	Texto	100
Independent_Variable	Texto	100
Direct_Context_I	Texto	100
Indirect_Context_I	Texto	100
EFFECT	Texto	50
Control_D	Texto	100
Magnitude_of_Difference_D	Texto	100
Dependent_Variable	Texto	100
Direct_Context_D	Texto	100
Indirect_Context_D	Texto	100
Original_Source_Hypothesis	Texto	255
Level_Support	Texto	50
Confidence_Support	Texto	50
Topic	Texto	255
OBS	Memorando	-

**Relacionamentos**

**PaperHypothesis**



Attributes:

RelationshipType:

Imposto

Um-para-muitos

Tabela: Paper

**Colunas**

Nome	Tipo	Tamanho
PaperID	Inteiro longo	4
PaperTitle	Texto	250
Publication	Texto	250
Year	Texto	50
HowPracticeWasApplied	Inteiro longo	4
HowResultsWasMeasured	Inteiro longo	4
HowResultsWasReported	Inteiro longo	4
WhoReportedTheResults	Inteiro longo	4
Context Description's Degree of Details	Inteiro longo	4
Abstract	Memorando	-
CompleteReference	Memorando	-

**Relacionamentos**

**PaperAnalysisXQualityCriteriaXPaper**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperContext Description**



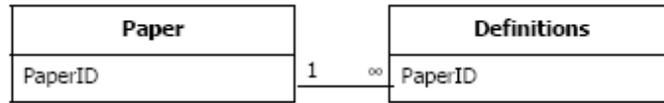
Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperContext\_General**



Attributes: Não imposto  
 RelationshipType: Um-para-muitos

**PaperDefinitions**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperHypothesis**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperConclusionsXPapers**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperAnalysisXPapers - Included Studies**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

**PaperPaperXAuthors**



Attributes: Imposto  
 RelationshipType: Um-para-muitos

Tabela: PaperXAuthors

**Colunas**

Nome	Tipo	Tamanho
AuthorID	Inteiro longo	4
PaperID	Inteiro longo	4

**Relacionamentos**

**AuthorsPaperXAuthors**



Attributes: Imposto  
RelationshipType: Um-para-muitos

**PaperPaperXAuthors**



Attributes: Imposto  
RelationshipType: Um-para-muitos

Tabela: QualityCriteria

**Colunas**

Nome	Tipo	Tamanho
CriteriaID	Inteiro longo	4
Criteria_Description	Texto	50
Criteria_PossibleReasoning	Memorando	-

**Relacionamentos**

**QualityCriteriaAnalysisXQualityCriteriaXPaper**



Attributes: Imposto  
RelationshipType: Um-para-muitos