

**Titulo:** Reutilização de Modelos de Requisitos por Analogia: Experimentação e Conclusões

**Autor:** Sérgio Felipe Zirbes

**Endereço:** Instituto de Informática/UFRGS, Av. Bento Gonçalves, 9500  
Bloco IV - Agronomia, Porto Alegre - RS - Brasil  
Caixa Postal 15064, CEP 91501-970  
Telefone (051) 330 5500 - Ramal 5175  
Fax +55(51)336.5576; e-mail: zirbes@inf.ufrgs.br

**Abstract:** This paper presents the conceptual framework, the organization, realization and analysis of a requirements reuse experiment. The reuse of requirements models is investigated in two different methodologies: according to the structured methodology, the modeling process is based on the use of Data Flow Diagrams; in the Object Orientation paradigm, Object Diagrams are used for modeling purposes. The research was achieved with the cooperation of 114 students/analysts, resulting in 175 series of Data Flow Diagrams and 23 series of Object Diagrams. Proper statistical analysis were conducted with these samples, in order to clarify questions about requirements reuse.

**Keywords:** Early Reuse, Systematic Reuse, Requirements Reuse, Reuse by Analogy, Object-Oriented Paradigm, Specification Reuse, Object Diagrams, Data Flow Diagrams.

## 1. INTRODUÇÃO

A reutilização sistemática [FRA 94] envolve uma série de atividades preliminares para criação da biblioteca de recursos reutilizáveis [PRI 91a]. A especificação dos requisitos de um novo sistema tem um ciclo de vida próprio [ARA 89, REU 89, WAT 91 e MAI 92a], envolvendo a eliciação destes requisitos, criação de seu modelo restrito, busca de modelos similares na biblioteca, seleção do modelo similar mais adequado e desenvolvimento do modelo do novo sistema. O pressuposto básico da reutilização, entretanto, é que esta sistemática *produza sistemas de maior qualidade e confiabilidade, de uma forma mais produtiva* [FRA 94]. A experimentação descrita neste artigo objetivou a investigação empírica deste pressuposto, centrando-se, pois, na atividade de construção dos modelos. Descreve-se a idealização, execução e análise de experimentos capazes de permitir a comparação de modelos de requisitos, referentes ao mesmo problema alvo, mas construídos em processos diferentes, com e sem reutilização. Os participantes dos experimentos foram alunos/analistas de cursos de especialização e mestrado na área de informática. Os problemas alvos foram definidos de forma a poderem ser modelados em período não superior a duas horas. Os modelos reutilizáveis foram representados em sua forma não abstrada, isto é, como modelos concretos, sendo que no experimento no paradigma estruturado a modelagem foi realizada com Diagramas de Fluxo de Dados (DFD's) [YOU 90] e no experimento no paradigma orientado a objeto, com Modelo de Objetos [COA 92]. A comparação viabilizou conclusões sobre eventuais ganhos de qualidade, completude e esforço com a reutilização, nos dois métodos de desenvolvimento de especificações de requisitos. Dois estudos distintos foram realizados: no primeiro, utilizando DFD's, alunos/analistas elaboraram 175 modelos, sendo 90 construídos desde o início, sem reutilização e 85 com reutilização de modelo similar; no segundo, utilizando-se Modelos de Objetos, obteve-se uma amostra menor, com 23 modelos, sendo 11 construídos sem e 12 com reutilização de modelo similar. *O objetivo da pesquisa realizada foi permitir a comparação de eficácia, eficiência e produtividade nas duas sistemáticas de construção de modelos, com e sem reutilização, no âmbito de cada um dos dois paradigmas.*

Este artigo, descrevendo a pesquisa, está estruturado em cinco seções. A esta introdução segue-se a análise do protocolo experimental e da estratégia pela qual foram regidos os experimentos de modelagem. Os experimentos de modelagem realizados com ferramentas inerentes ao paradigma estruturado (DFD's) e ao paradigma orientado a objeto (Modelos de Objetos) são descritos numa terceira fase. Características dos participantes, descrição das experiências e do material utilizado em sua execução e critérios de avaliação dos modelos são, também, aspectos considerados neste item. Na parte central do artigo são apresentadas as conclusões do trabalho. De forma criteriosa, busca-se extrapolar os resultados obtidos na pesquisa, aplicando-os ao processo genérico de reutilização de modelos de requisitos de sistemas. Os efeitos do grau de similaridade entre o problema alvo e o problema reutilizável em relação à qualidade e completude dos modelos resultantes são realçados. Aspectos que devem ser ainda pesquisados e idéias quanto à continuação futura do trabalho constituem a seção final do trabalho.

## 2. PLANO DE EXECUÇÃO DA PESQUISA

A validação de uma teoria através do uso de experiências, passíveis de serem repetidas por outros pesquisadores, é um dos principais meios utilizados no desenvolvimento do conhecimento científico. Entretanto, apesar deste fato ser amplamente reconhecido, freqüentemente teorias são difundidas e aceitas sem que tenham sido comprovadas. Esta constatação é ainda mais verdadeira quando trata-se de teorias inerentemente difíceis de ser validadas ou ainda com aquelas que, aparentemente passíveis de aceitação intuitiva, sugerem ser desnecessária sua comprovação [LEW 91]. A validação experimental largamente utilizada

em Engenharia, Física, Biologia e Medicina, raramente é utilizada nos problemas relacionados com a Engenharia de Software

Os estudos empíricos realizados por Maiden e Sutcliffe [MAI 92 e MAI 92a], a partir de 1990, entretanto, constituem-se em importante paradigma para análises deste tipo. Os experimentos realizados por Maiden, além de avaliar o processo de reutilização por analogia, buscam identificar as facilidades necessárias a um mecanismo de recuperação de especificações. Abrangendo escopo similar ao estudado por Maiden, nossa pesquisa analisou os aspectos de produtividade envolvidos no processo de elaboração de modelos de requisitos de sistemas através da analogia com modelos similares previamente construídos [ZIR 93].

Um dos principais itens do plano experimental é o *tipo de projeto de pesquisa a ser adotado*. Dentre os três projetos experimentais propriamente ditos apresentados por Campbell e Stanley em [CAM 69], um dos que menos influência sofre em relação aos fatores que podem comprometer sua validade, é o denominado *projeto de pesquisa com grupo de controle e pós-teste*. Neste tipo de plano, busca-se comparar o comportamento observado em um grupo de indivíduos que não sofreu determinada ação externa (o grupo de controle) com o comportamento de outro grupo da mesma população que tenha sofrido a ação. Com base neste esquema, os experimentos foram realizados mediante a utilização de turmas de estudantes de especialização e pós-graduação, de acordo com os seguintes passos:

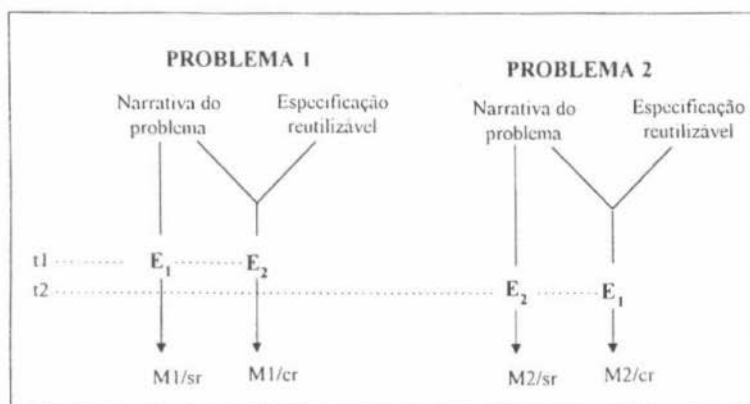


Figura 2: Esquema dos experimentos aplicados aos dois paradigmas

- E1 - No momento t1, sem reutilização, a partir da narrativa do problema 1, elaborou o modelo do problema alvo M1/sr. No momento t2, a partir da narrativa do problema 2, agora com reutilização, construiu o modelo do segundo problema-alvo, M2/cr.
- E2 - No momento t1, desenvolveu o modelo do problema-alvo M1/cr com reutilização de modelo similar. No momento t2, construiu o modelo do segundo problema alvo, M2/sr, agora sem qualquer reutilização.

No final dos experimentos, em cada um dos dois paradigmas foram construídos 4 modelos, sendo dois deles sem reutilização e dois com reutilização. Ambas as equipes, E1 e E2 tiveram a oportunidade de exercitar a reutilização.

### 3. DESCRIÇÃO DOS EXPERIMENTOS DE MODELAGEM

Após ter sido definido o protocolo experimental e a estratégia a ser seguida na pesquisa, é necessário descrever como os dados correspondentes foram organizados de forma a possibilitar a análises referidas no objetivo da pesquisa. Como já foi mencionado, a reutilização de modelos foi investigada em dois paradigmas diferentes: estruturado, com DFD's e orientado a objeto, com Modelos de Objetos.

#### 3.1 Os experimentos de modelagem

As experiências de modelagem com DFD's tiveram a participação de quatro turmas distintas de alunos/analistas, sendo que 20 % deles podem ser considerados experientes em análise, 44% de experiência média e 36% inexperientes. Em sua maioria, os participantes dos experimentos no paradigma estruturado frequentaram curso formal de graduação em informática (74 %). Os demais (26 %), complementaram sua formação com outros cursos na área.

Os dois experimentos no paradigma orientado a objeto foram realizados pelos 13 alunos da turma de 1994 do Curso de Especialização em Análise de Sistemas da Pontifícia Universidade Católica do Rio Grande do Sul - PUC/RS. Quando da realização dos experimentos os alunos/analistas tinham já recebido o conteúdo programático da disciplina, no que diz respeito aos fundamentos teóricos da análise orientada a objeto, de acordo com a metodologia de Coad e Yourdon [COA 92]. Havia ainda realizado uma série de exercícios práticos de modelagem. Embora não tivessem experiência prévia em análise orientada a objetos, os alunos analistas em sua maioria tinham uma experiência média em análise. Por outro lado, 7 participantes se graduaram em cursos formais da área de informática, enquanto que 6 realizaram cursos em outras áreas.

Cada turma de participantes foi dividida em duas equipes, E1 e E2. Cada uma das equipes desenvolveu dois modelos de problemas diferentes, sendo o primeiro construído desde o início e o segundo desenvolvido com reutilização de diagramas de aplicação similar:

##### a. Modelagem com DFD's

- Exp. 1 - problema alvo: controle de locadora de veículos  
problema reutilizável: controle de locadora de fitas de vídeo
- Exp. 2 - problema alvo: controle de alocação de leitos em hospitais  
- problema reutilizável: controle e emissão de passagens aéreas

##### b. Modelagem com Modelo de Objetos

- Exp. 1 - problema alvo: controle de bibliotecas  
- problema reutilizável: reservas de passagens aéreas
- Exp. 2 - problema alvo: controle e emissão de carteiras de motorista  
- problema reutilizável: controle e registro de propriedade de veículos

A adoção deste esquema permitiu a análise de 175 modelos no paradigma estruturado (90 construídos sem reutilização e 85 com aproveitamento de modelo similar), e de 23 modelos no paradigma orientado a objeto (11 modelados sem reutilização e 12 com aproveitamento de modelo similar).

#### 3.2 Descrição do material distribuído nos experimentos

Para que os experimentos pudessem ser desenvolvido foram distribuídos, no início de cada experiência, cadernos de instruções, contendo os seguintes itens, devidamente adaptados a cada tipo de modelagem:

1. Resumo de alguns aspectos da metodologia e exemplos de modelos.

2. Descrição das etapas da construção de modelos de requisitos (com e sem reutilização).
3. Descrição narrativa do problema alvo.
4. Descrição narrativa e modelos para reutilização (apenas para as equipes que iriam reutilizar).
5. Tabela para apropriação dos tempos envolvidos nas diversas etapas da modelagem dos requisitos.
6. Folhas em branco para modelagem
7. Espaço para colocação de observações e anotações que o participante desejasse registrar.

### 3.3 Critérios e método de correção dos modelos desenvolvidos.

A completude e a adequação do modelo à realidade foram os principais aspectos considerados na avaliação dos modelos desenvolvidos pelos alunos/analistas. Para que a avaliação fosse isenta e objetiva, foram estabelecidos critérios de correção:

#### ESQUEMA DE CORREÇÃO DOS DIAGRAMAS

1. Modelagem com DFD's			
<b>Diagrama Contextual</b>	<b>(35)</b>	<b>Diagrama de nível 0</b>	<b>(65)</b>
1. Falha total no desenho	-35	1. Falha total no desenho	-65
<b>Erros de Sintaxe</b>	<b>(15)</b>	<b>Erros de Sintaxe</b>	<b>(25)</b>
2. Ausência de nomes de fluxos	-5	1. Ausência de nomes de fluxos	-5
3. Mau uso de outros símbolos	-5	2. Mau uso de outros símbolos	-10
4. Regras de constr. e visibilidade	-5	3. Regras de constr. e visibilidade	-10
<b>Erros de Semântica</b>	<b>(20)</b>	<b>Erros de Semântica</b>	<b>(40)</b>
5. Erro na identif. de terminadores	-10	4. Erro na identif. dos processos	-15
6. Falha na identif. de interfaces	-10	5. Identif. de arquivos e interfaces	-15
		6. Desbalanceamento	-10

2. Modelagem com Objetos	
<b>Diagrama Orientado a Objeto</b>	<b>(100)</b>
1. Falha total no modelo	-100
<b>Modelo de Dados</b>	<b>(60)</b>
2. Identif. de classes e objetos	-20
3. Identificação das estruturas	-20
4. Atributos	-15
5. Assuntos	-5
<b>Modelo Dinâmico</b>	<b>(40)</b>
6. Identificação dos serviços	-20
7. Identif. conexões de mensagem	-20

Os pontos expressos ao lado de cada tipo de erro correspondem ao número máximo de pontos que o aluno/analista pode *perder* devido a este erro. Assim, uma avaliação dos modelos igual a 100, equivale a uma modelagem sem erros. Ainda exemplificando: a um conjunto de modelos onde o diagrama contextual estivesse correto e onde no diagrama de nível 0 houvesse erros semânticos, correspondentes a ausência do nome de alguns fluxos (-5) e erro na identificação dos principais processos (-15), a avaliação final seria de 80 ((100 - (5+15)).

### 3.4 Apropriação dos dados dos experimentos: caracterização das variáveis

Relativamente a cada um dos experimentos, foram levantados três conjuntos de dados: caracterização e experiência dos participantes, avaliação dos modelos construídos e tempos necessários para a execução das diversas etapas da modelagem. Mais detalhes sobre estas variáveis podem ser obtidos em [ZIR 95].

## 4. ANÁLISE DA PESQUISA E CONCLUSÕES

O objetivo da análise dos resultados dos experimentos 1 e 2, tanto na modelagem com DFD's como na modelagem com Diagrama de Objetos é o de comparar a qualidade e completude dos modelos, e a produtividade dos participantes nos dois processos de modelagem, com e sem reutilização de modelos por analogia. Os resultados desta comparação serão então analisados e confrontados com as características de cada processo de modelagem, visando-se *explicar e justificar* eventuais vantagens de um processo sobre o outro. Dificuldades e problemas inerentes ao processo de reutilização deverão ser apontados, e, finalmente, são enumerados uma série de aspectos da reutilização sistemática que ainda requerem pesquisa adicional, delineando-se assim o roteiro para continuação do presente trabalho.

### 4.1 Análise crítica dos resultados dos experimentos com DFD's

Existem algumas diferenças básicas entre os experimentos 1 e 2. O experimento 2 foi realizado sempre, em todas as turmas, um determinado período de tempo após o experimento-1. Estes intervalos foram diferentes para cada turma, variando de um dia a uma semana. Já foi ressaltado que os problemas a serem modelados no experimento-2 foram diferentes daqueles utilizados no experimento-1. Convém lembrar, ainda, que os participantes do experimento-2 também participaram do experimento-1, embora com papéis invertidos (os participantes que reutilizaram no experimento-1 modelaram sem reutilização no experimento-2 e vice-versa). Dois aspectos, portanto, diferenciam os experimentos 1 e 2:

**a. a questão do aprendizado** - quando os participantes realizaram o experimento-2, estavam já mais familiarizados com as técnicas de modelagem de diagramas e talvez de reutilização de diagramas do que no experimento-1. Sob este aspecto, os dados do experimento-1 refletem melhor os objetivos da pesquisa que os do experimento-2 ou da amostra com os dois experimentos em conjunto.

**b. a questão dos problemas serem diferentes** - a diferença eventual na complexidade do problema alvo entre os dois experimentos conduziu a resultados diferentes. Este efeito foi aumentado em função da diferença no grau de similaridade dos problemas cujos modelos foram reutilizados nos dois experimentos. Durante a análise da pesquisa, serão feitas maiores referências a estes dois aspectos.

#### 4.1.1 Variações de AF em função do tipo de modelagem

Uma das variáveis mais importantes no contexto da experimentação proposta pela tese é AF, avaliação final do modelo, que mede o grau de completude e correção. Quanto mais alta for a média de AF para os modelos de uma determinada amostra, maior o grau de completude e correção destes modelos. A tabela 1 apresenta dados sobre AF em três amostras distintas: experimento-1, experimento-2 e conjunto dos dois experimentos com DFD's.

Experimentos com Diagramas de Fluxo de Dados				
Variação de média, desvio padrão, valores mínimo e máximo de AF				
Comparação dos dados obtidos com e sem reutilização				
		Modelagem		
Experimento	Dados	S/reut.	C/reut.	Variação
1	Contagem	51	50	
	Média de AF	64,51	84,80	31,45
	DesvPad de AF	20,89	9,95	
2	Contagem	39	35	
	Média de AF	64,49	75,00	16,30
	DesvPad de AF	16,81	9,39	
Total Contagem		90	85	
Total Média de AF		64,50	80,76	25,22
Total DesvPad de AF		19,12	10,82	

Tabela 1: Variação de AF de acordo com o tipo de modelagem

**a. experimento-1** - A tabela 1 mostra claramente que os diagramas construídos mediante reutilização de modelos análogos são **mais completos e corretos** do que os diagramas construídos desde o início, sem reutilização. A variação da média de AF, obtida no experimento-1, é de 31,45%. A medida do desvio padrão, reduzida à metade com a reutilização, mostra que neste processo de modelagem os resultados foram mais homogêneos.

**b. experimento-2** - Todas as conclusões obtidas pela análise do experimento-1 são válidas para o experimento-2. Há, entretanto, uma diferença fundamental: **os ganhos em completude e correção caem à metade em relação aos do experimento-1: 16,3%**. O próximo item analisa justamente as razões para esta diferença.

**c. diferença no ganho de correção e amplitude entre os dois experimentos** - A tabela 1 mostra que as médias de AF, sem reutilização, são praticamente iguais nos dois experimentos (64,51 e 64,49). Este fato demonstra que **os dois problemas alvo tem grau de complexidade semelhante**. Considerando-se a questão do aprendizado já referida, a igualdade das médias de AF nos dois experimentos pode sugerir que o segundo experimento é um pouco mais complexo que o primeiro. Assim, uma pequena diminuição do ganho em correção e amplitude no experimento-2 era de certa forma esperada. Entretanto, a variação da média de AF com e sem reutilização no experimento-2 foi mais ou menos a metade daquela obtida com o experimento-1, 16,3%. Uma variação desta ordem não pode ser atribuída apenas a uma pequena diferença de complexidade (aliás compensada pela questão do aprendizado)

O que realmente justifica o ganho inferior é a **diferença no grau de similaridade entre os problemas alvo e os problemas cujos modelos foram reutilizados**. No primeiro caso, os dois sistemas eram muito semelhantes (controle de locação de fitas de vídeo e de automóveis).

Para que se obtivesse o modelo final no experimento-1, bastava substituir os nomes dos terminadores e fluxos no diagrama contextual reutilizável. Já no diagrama de nível 0, era possível utilizar-se os mesmos processos bastando renomeá-los. As duas especificações narrativas tinham a mesma estrutura, ressaltando a analogia.

No segundo experimento, entretanto, a similaridade embora considerável, era bem menor. Neste caso, o problema alvo era um sistema de controle de alocação de leitos em hospitais e o problema reutilizável era um sistema de controle de reserva de passagens aéreas. As duas descrições narrativas foram apresentadas numa forma não-estruturada. Uma das diferenças básicas entre os dois problemas era a existência, no problema do controle de leitos hospitalares, de um arquivo denominado *prontuário* onde devem ser armazenados todos os

hospitalares, de um arquivo denominado *prontuário* onde devem ser armazenados todos os dados clínicos do paciente. Além deste arquivo deveria ser previsto outro, com os dados financeiros/administrativos. Vários alunos analistas, ao reutilizar os diagramas do problema referente ao controle das passagens aéreas não foram capazes de verificar a existência adicional do *prontuário*. Simplesmente, num processo típico de *cópia*, já referenciado na bibliografia [NOV 88, CHI 89 e MAI 92], modelaram a solução com um arquivo único, misturando informações de caráter financeiro/administrativo e informações clínicas. Além disso, a ênfase no problema alvo era o *trâmite de documentação e o processo de pagamento*, ao passo que no problema reutilizável era a *reserva das passagens*.

#### 4.1.2 Variações dos tipos de erros em função do tipo de modelagem

Outro objetivo importante da pesquisa foi verificar se a reutilização de modelos análogos inibia mais particularmente alguns tipos específicos de erros. O esquema de avaliação dos modelos construídos durante os experimentos, adotou a seguinte taxinomia: perda de pontos por erros sintáticos (P1) e semânticos (P2) no diagrama contextual e perda de pontos por erros sintáticos (P3) e semânticos (P4) no diagrama de nível 0. O total de pontos perdidos devido a erros, (PT) é simplesmente a soma de P1, P2, P3 e P4.

Experimentos com Diagramas de Fluxo de Dados				
Variação dos vários tipos de erros, P1, P2, P3, P4 e PT				
Comparação dos dados obtidos com e sem reutilização				
		Modelagem		
Experimento	Dados	S/reut.	C/reut.	Variação
1	Contagem	51	50	
	Média de P1	2,75	0,10	96,36
	Média de P2	7,59	4,04	46,76
	Média de P3	11,08	2,00	81,95
	Média de P4	14,08	9,06	35,65
	Média de PT	35,49	15,20	57,17
2	Contagem	39	35	
	Média de P1	0,90	0,86	4,49
	Média de P2	12,69	9,29	26,84
	Média de P3	6,15	4,14	32,68
	Média de P4	15,77	10,71	32,06
	Média de PT	35,51	25,00	29,60
Total Contagem		90	85	
Total Média de P1		1,94	0,41	78,82
Total Média de P2		9,80	6,20	36,73
Total Média de P3		8,94	2,88	67,77
Total Média de P4		14,81	9,74	34,23
Total Média de PT		35,50	19,24	45,82

Tabela 2: Variação dos erros com o tipo de modelagem

A tabela 2 apresenta os dados relativos aos erros na modelagem com e sem reutilização, no experimento-1, no experimento-2 e na amostra total dos dois experimentos em conjunto.



**a. experimento-1** - reforçando-se o que já se imaginava, a redução dos diversos tipos de erros na modelagem com reutilização é realmente considerável. Em média, os pontos perdidos devido a erros são reduzidos a menos da metade (de 35,49 a 15,20). Mais do que isto, os erros sintáticos são quase eliminados, com reduções da ordem de até 93 % no diagrama contextual. Já com os erros semânticos, ocorrem reduções da ordem de 47 %, também no diagrama contextual. Deduz-se a partir daí que a reutilização de modelos por analogia, reduz mais os erros sintáticos que os erros semânticos, servindo para ensinar os alunos/analistas menos experientes a construir modelos sintaticamente mais corretos.

**b. experimento-2** - os dados do experimento-2, embora corroborem o fato de que a reutilização induz de fato à construção de modelos mais corretos e completos, não confirma serem os erros sintáticos os que maior redução sofrem.

**c. diferença entre a diminuição de erros nos dois experimentos** - É importante ressaltar que no segundo experimento, os erros sintáticos, já na modelagem sem reutilização foram menos frequentes. Isto pode ser explicado pela *questão do aprendizado*. Já os erros semânticos aumentaram no segundo experimento, provavelmente devido ao *menor grau de similaridade existente entre o problema alvo e o problema reutilizável*, fato já discutido.

#### 4.1.3 Variações dos tempos em função do tipo de modelagem

A questão dos tempos gastos por atividade nos dois processos de modelagem com e sem reutilização, é de fundamental importância. Esperávamos que aos eventuais benefícios da reutilização fosse contrapor-se um aumento no esforço de modelagem, devido às atividades adicionais inerentes ao segundo processo: afinal, é preciso ler e entender a especificação reutilizável antes de se tornar possível a reutilização. Para avaliar corretamente estas diferenças no esforço de modelagem, apropriamos, durante os experimentos, os tempos gastos pelos alunos/analistas em cada atividade. Isto foi feito através de uma planilha inserida no material distribuído aos participantes, que as iam preenchendo a medida que realizavam as diversas atividades da modelagem. Além de verificar o tempo adicional gasto para ler e compreender a especificação análoga, buscava-se também verificar a variação do esforço que a reutilização determinaria sobre as tarefas comuns aos dois processos. O esquema da planilha permitiu a apropriação do tempo para ler a especificação narrativa do problema alvo (T1), tempo para ler e compreender a especificação reutilizável (T2), quando da modelagem com reutilização, tempo para construir o diagrama contextual (T3), tempo para construir o diagrama de nível 0 (T4), e tempo para revisão final dos dois diagramas (T5). Mais tarde, na consolidação das tabelas para análise estatística, foram criadas as variáveis TA (tempo administrativo = soma de T1, T2 e T5) e TC (tempo de construção = soma de T3 e T4). A tabela 3 fornece os dados relativos aos tempos consumidos em cada etapa nos dois processos de modelagem, nos dois experimentos.

**a. experimento 1** - observa-se, em primeiro lugar, que o tempo total gasto na construção dos diagramas nos dois tipos de modelagem **difere muito pouco**. O tempo que mais redução sofreu (26,15) foi justamente a tarefa mais difícil, que é a construção do diagrama de nível 0. Mas, a conclusão mais importante e a de que *a redução nos tempos de construção dos diagramas compensou a pequena variação nos tempos administrativos* (devida à necessidade de uma análise adicional da aplicação reutilizável...), permitindo que a média do tempo total, variável TT diminuisse ainda na modelagem com reutilização.

**b. experimento-2** - os resultados do experimento-2 mostram um comportamento diferente, na medida em que o tempo adicional devido a leitura e entendimento da especificação reutilizável não é mais compensado por uma diminuição no tempo de construção dos diagramas.

**c. diferença na distribuição dos tempos nos experimento 1 e 2** - como já foi referido, uma complexidade de um pouco maior do problema e um grau de similaridade menor entre o problema

alvo e o problema reutilizável, fizeram com que a construção dos diagramas no experimento-2 fosse sensivelmente mais difícil do que no experimento-1. Isto fez com que a redução de TC fosse menor, causando diferença dos resultados em relação ao experimento-1.

Experimentos com Diagramas de Fluxo de Dados				
Variação dos tempos nas atividades, T3, T4, TA, TC e TT				
Comparação dos dados obtidos com e sem reutilização				
		Modelagem		
Experimento	Dados	S/reut	C/reut	Variação
1	Contagem	51	50	
	Média de T3	12,16	11,08	8,86
	Média de T4	23,24	17,16	26,15
	Média de TA	33,00	33,14	-0,42
	Média de TC	35,39	28,24	20,21
	Média de TT	68,39	61,38	10,25
2	Contagem	39	35	
	Média de T3	11,95	12,34	-3,30
	Média de T4	18,13	15,09	16,78
	Média de TA	24,49	27,46	-12,13
	Média de TC	30,08	27,43	8,81
	Média de TT	54,56	54,89	-0,59
Total Contagem		90	85	
Total Média de T3		12,07	11,60	3,87
Total Média de T4		21,02	16,31	22,44
Total Média de TA		29,31	30,80	-5,08
Total Média de TC		33,09	27,91	15,66
Total Média de TT		62,40	58,71	5,92

Tabela 3: Variação dos tempos em função do tipo de modelagem

A análise realizada dos resultados dos experimentos no paradigma estruturado, com modelagem através de DFD's, permite a seguinte constatação: **mediante reutilização de modelos análogos, foi possível a construção de diagramas mais corretos e completos, sem aumento considerável de esforço.**

#### 4.2 Análise crítica dos resultados dos experimentos com Modelos de Objetos

O objetivo da pesquisa no que se refere aos dois experimentos com os diagramas de objetos, de acordo com a metodologia proposta por Coad & Yourdon [COA 92], foi o de verificar se os resultados obtidos na reutilização de DFD's poderiam ser estendidos também a outras ferramentas de modelagem. A escolha dos diagramas de objetos, como segundo paradigma de modelagem se deu em razão de que estes modelos agregam aspectos do comportamento dinâmico do sistema aos aspectos estáticos modelados por outras ferramentas (modelo entidade-relacionamento, por exemplo). Mesmo com uma pequena amostra, composta de 23 modelos, algumas considerações importantes podem ser formuladas a partir da pesquisa no paradigma orientado a objeto.

Os experimentos no POO foram feitos com o mesmo protocolo experimental e a mesma estratégia utilizada quando dos experimentos com DFD's. As considerações sobre validade aplicam-se novamente, já que o experimento-2 foi realizado com um problema levemente mais

complicado e com reutilização de modelo com grau de similaridade menor do que no experimento-1. Novamente o experimento-1 deve ser considerado mais representativo do que o experimento-2, especialmente em função da *questão do aprendizado* já discutida na análise dos experimentos com DFD's.

#### 4.2.1 Variações de AF em função do tipo de modelagem

A variação da completude e correção dos diagramas de objetos em decorrência da reutilização foi bastante diferente nos dois experimentos. A análise realizada sobre a tabela 4 busca justificar este resultado.

No experimento-1 a variação da média de AF foi bastante alta, na ordem de 70 %. O teste t, neste caso, elimina a hipótese da casualidade, o que implica ser possível admitir que a reutilização proporcionou uma grande melhoria na completude e correção dos diagramas de objetos.

Também no segundo experimento com diagrama de objetos a qualidade dos modelos aumentou com a reutilização. Entretanto, o ganho, neste caso, foi bem menor: uma variação de apenas 13,87 % em AF. No próximo item analisamos as causas deste comportamento.

Experimentos com Modelos de Objetos				
Variação da média, desvio padrão, valores mínimo e máximo de AF				
Comparação dos dados obtidos com e sem reutilização				
		Modelagem		
Experimento	Dados	S/reut.	C/reut.	Variação
1	Contagem	5	5	
	Média de AF	42,20	72,00	70,62
	DesvPad de AF	6,50	10,44	
2	Contagem	6	7	
	Média de AF	56,83	64,71	13,87
	DesvPad de AF	4,79	7,70	
Total Contagem		11	12	
Total Média de AF		50,18	67,75	35,01
Total DesvPad de AF		9,31	9,27	

Tabela 4: Variação de AF de acordo com o tipo de modelagem

Na tabela 4 verificamos que a média de AF na modelagem sem reutilização no experimento-2 (56,83) é superior a mesma média no experimento-1 (42,20). Ao analisarmos as especificações narrativas dos problemas alvo nos experimentos 1 e 2 verificamos que os requisitos do *sistema de controle de bibliotecas* e do *sistema de controle e emissão de carteiras de motorista* são igualmente simples. A diferença na média de AF, neste caso, deve ser atribuída ao *efeito aprendizado*. Por outro lado, na análise das médias de AF, na modelagem com reutilização, verificamos que os participantes no segundo experimento obtiveram um índice menor (64,71) do que o obtido pelo grupo que reutilizou no primeiro experimento (72,00). Trata-se, novamente, do efeito da diferença do grau de similaridade entre problema alvo e problema reutilizável nos dois experimentos. No experimento-1, o *sistema de controle de biblioteca* e o sistema de controle de locações de fitas de vídeo são bastante similares (o que varia de fato são os nomes dos objetos e seus atributos, mas não os procedimentos ou *serviços* relacionados). O mesmo já não ocorre com o *sistema de emissão de títulos de propriedade* e o *sistema de emissão e controle de carteiras de motorista*

utilizados no experimento-2. O problema alvo, neste caso, envolve um conjunto de procedimentos relativos à organização e apropriação de resultados do curso teórico de direção, sem similar no problema reutilizável. Repete-se o que já havia ocorrido na modelagem com DFD's.

#### 4.2.2 Variações dos tipos de erros em função do tipo de modelagem

Neste ponto da pesquisa busca-se evidenciar uma eventual diferença no efeito da reutilização sobre os aspectos dinâmicos e estáticos do diagrama de objetos. Na tabela 5, a variável ME (de modelo estático), é a soma dos pontos perdidos pelos participantes na modelagem de *classes e objetos* (CO), *estruturas* (ES), *atributos* (AT) e *assuntos* (AS). Já MD (de modelo dinâmico), é a soma dos pontos perdidos na modelagem de *serviços* (SE) e *conexões de mensagem* (CM).

Experimentos com Modelos de Objetos				
Variação dos erros associados aos modelos estático e dinâmico				
Comparação dos dados obtidos com e sem reutilização				
		Modelagem		
Experimento	Dados	S/reut.	C/reut.	Variação
1	Contagem	5	5	
	Média de ME	26,80	15,40	42,54
	Média de MD	31,00	12,60	59,35
2	Contagem	6	7	
	Média de ME	19,50	13,43	31,14
	Média de MD	23,67	21,86	7,65
Total Contagem		11	12	
Total Média de ME		22,82	14,25	37,55
Total Média de MD		27,00	18,00	33,33

Tabela 5: Variação dos erros com o tipo de modelagem

No experimento-1, onde o modelo do problema alvo e o modelo do problema reutilizável eram bastante similares, houve maior redução percentual na média de MD (59,35 %) do que na média de ME (42,54 %). Já no segundo experimento, os resultados são diferentes, com a redução da média de ME ficando em 31,14 % e a de MD em apenas 7,65 %. A questão do grau de similaridade entre problema alvo e problema reutilizável parece justificar a inversão ocorrida.

#### 4.2.3 Variações dos tempos em função do tipo de modelagem

Relativamente aos tempos gastos pelos participantes nas diversas atividades de modelagem com diagrama de objetos segundo os dois processos, com e sem reutilização, esperava-se a repetição do que ocorreu com os experimentos no paradigma estruturado: apenas uma pequena variação no tempo total da modelagem. A tabela 6, de fato confirma a tendência.

### 5. A CONTRIBUIÇÃO DA PESQUISA

As principais conclusões resultados e contribuições da pesquisa serão apresentados em duas seções, de acordo com a origem das informações que lhes dão sustentação: resultados da análise estatística e observações colhidas durante os experimentos e nos questionários preenchidos por todos os alunos/analistas que participaram dos experimentos de modelagem reutilizando modelos análogos.

Experimentos com Modelos de Objetos				
Variação dos tempos nas atividades durante a modelagem				
Comparação dos dados obtidos com e sem reutilização				
		Modelagem		
Experimento	Dados	S/reut.	C/reut.	Variação
1	Contagem	5	5	
	Média de TA	20,40	26,20	-28,43
	Média de TC	38,40	37,80	1,56
	Média de TT	58,80	64,00	-8,84
2	Contagem	6	7	
	Média de TA	14,67	29,14	-98,70
	Média de TC	49,67	32,57	34,42
	Média de TT	64,33	61,71	4,07
Total Contagem		11	12	
Total Média de TA		17,27	27,92	-61,62
Total Média de TC		44,55	34,75	21,99
Total Média de TT		61,82	62,67	-1,37

Tabela 6: Variação dos tempos em função do tipo de modelagem

### 5.1 Resultados da análise estatística.

A realização de quatro experimentos, sendo dois com DFD's e dois com Diagramas de Objetos, todos visando comparar resultados da modelagem com e sem reutilização, evidenciou, de acordo com o esperado, uma série de conclusões sobre o processo de modelagem de requisitos:

**a. completude e correção dos modelos** - Sem dúvida, a principal conclusão que a análise dos experimentos proporcionou foi a de que a modelagem de requisitos de sistemas de informação através da reutilização de modelos de requisitos de sistemas análogos produz resultados mais completos e corretos do que a modelagem tradicional, sem reutilização. A análise do experimento-1 mostrou que a média de AF nos modelos construídos com reutilização foi na ordem de 30 % mais alta do que a média de AF alcançada nos demais modelos. Entre as diversas turmas que participaram do experimento, o aumento da média de AF na modelagem com reutilização variou de 50 a 20 %. O teste t, realizado para verificar se estes resultados poderiam ser atribuídos ao acaso, negou esta hipótese, permitindo atribuir o ganho em completude e correção fosse atribuído ao novo paradigma de modelagem.

**b. efeitos da reutilização sobre os erros de modelagem** - os erros nos experimentos com DFD's foram diferenciados em sintáticos e semânticos. O experimento-1 permitiu concluir que na modelagem com reutilização ocorre uma maior redução dos erros relacionados a aspectos sintáticos da modelagem de DFD's (no modelo de nível 0, os erros sintáticos foram reduzidos em 82 % e os erros semânticos em 35 %). Novamente a hipótese do acaso foi excluída em função dos resultados do teste t.

**c. tempos gastos nas diversas etapas da modelagem** - os tempos gastos na leitura das especificações e revisões dos modelos, considerados tempos administrativos, não variaram muito nos dois tipos de modelagem, mesmo considerando-se o tempo adicional gasto com a especificação reutilizável. Por outro lado, os tempos efetivamente gastos para *construir* os diagramas contextual e nível 0 diminuíram na ordem de 20 % na modelagem com reutilização. As médias de TT, tempo total gasto pelos participantes nos dois tipos de modelagem foram semelhantes (diferença na ordem de 10 %). Conclui-se, aqui, que a diminuição dos tempos de

construção dos modelos, basta para compensar o pequeno aumento do tempo administrativo devido a necessidade de ler e compreender o modelo reutilizável.

**d. experiência dos participantes** - a expectativa de que houvesse um benefício maior da reutilização entre os participantes de maior experiência, não pode ser comprovada. Embora os dados do experimento-1 mostrassem uma tendência nesta direção, o teste  $t$  não descartou a hipótese da casualidade para a variação das médias entre os participantes de menor e maior experiência.

**e. a questão do benefício da reutilização** - exclusivamente no âmbito dos experimentos, onde já haviam sido selecionados e postos à disposição dos participantes os modelos reutilizáveis, pode-se afirmar que na modelagem realizada com reutilização ocorreu um *benefício relativo* [KAR 89]. Mesmo aceitando-se que a média do tempo total na construção dos modelos com reutilização é levemente maior que a média correspondente na modelagem sem reutilização, as vantagens diretas (advindas da maior correção e completude dos modelos) e as indiretas (gastos menores com manutenção) certamente compensam o esforço adicional.

**f. os resultados dos experimentos com diagramas de objetos** - fazendo-se as ressalvas necessárias quanto ao tamanho da amostra e falta de experiência dos participantes com análise orientada a objeto, é possível afirmar que os resultados destes experimentos tendem a confirmar as conclusões obtidas com a modelagem através de DFD's. Novamente os modelos construídos com reutilização foram mais completos e corretos e mais uma vez o aumento dos tempos administrativos foi compensado pela diminuição dos tempos gastos na construção dos modelos. Os resultados dos experimentos no POO e no Paradigma Estruturado permitem supor que com outros modelos, a exemplo dos Diagramas de Entidade Relacionamento, a reutilização tenha resultado semelhante.

**g. o grau de similaridade dos modelos** - ficou demonstrado também, a existência de uma série de pré-requisitos para que o benefício da reutilização seja expressivo. A realização do experimento-2 em ambos os paradigmas deixou claro que o grau de similaridade entre o modelo do problema alvo e o do problema reutilizável é fator decisivo no sucesso ou insucesso da modelagem com reutilização. Os resultados foram bastante conclusivos: quanto menor for o grau de similaridade entre o novo problema e o problema reutilizável, menores serão as diferenças na correção e completude entre os métodos de modelagem com e sem reutilização. De maneira contrária, os tempos na modelagem com reutilização aumentam em relação aos tempos na modelagem desde o início, sem reutilização, a medida que a similaridade diminui. Em outras palavras, **os benefícios da reutilização caem drasticamente com a diminuição do grau de similaridade**.

## 5.2 Observações e impressões colhidas durante os experimentos

A pesquisa desenvolvida no âmbito do Paradigma Estruturado, com a modelagem sendo feita com DFD's, envolveu, em média, cerca de 20 horas de contato em aula com cada uma das quatro turmas *Sinte193*, *Sinte194*, *Puc* e *Join*. Na preparação e realização dos dois experimentos no Paradigma Orientado a Objeto, foram gastos mais 25 horas com uma nova turma de curso de especialização da Puc. Foram, portanto, mais de 100 horas de contato direto com 114 analistas, a maioria deles empregados nas mais variadas empresas da Grande Porto Alegre, e de Joinville em Santa Catarina, caracterizando sem dúvida, uma boa amostra do universo de analistas de nossas empresas de desenvolvimento. Das inúmeras informações obtidas durante as aulas prévias, da observação direta realizada durante os exercícios e experimentos, dos contatos informais e do questionário de avaliação preenchido pelos participantes após a modelagem com reutilização, surgiram uma série de considerações importantes que serão discutidas a seguir.

**a. metodologias de desenvolvimento** - a principal dificuldade observada no que se refere à reutilização foi a pouca familiaridade **prática** dos alunos/analistas com metodologias de desenvolvimento. Os sistemas de informação, na maioria das empresas, ainda são desenvolvidos pelos métodos tradicionais. Diante deste quadro, a modelagem através da reutilização parece bastante distante da realidade de suas atividades atuais. A expectativa prévia dos participantes, portanto, era a de que iriam participar de uma experiência acadêmica sem perspectiva, para eles, de utilização num futuro próximo.

**b. a impressão dos alunos/analistas após os experimentos** - a medida que, durante as aulas prévias, os alunos/analistas iam se familiarizando com a idéia de reutilização por analogia, seu interesse também crescia. No momento da divisão das turmas, no experimento-1, a frustração daqueles que deveriam modelar desde o início, sem reutilização só era neutralizada com a informação de que no experimento seguinte, com outro problema, os papéis seriam invertidos. A avaliação dos participantes após terem participado de um experimento com reutilização foi a melhor possível: todos os 114 responderam afirmativamente à pergunta sobre as possibilidades práticas do processo de modelagem com reutilização por analogia.

**c. a avaliação do processo de reutilização proposto** - as etapas do processo de reutilização que foram seguidas pelos participantes dos experimentos, previamente validadas na disciplina ministrada no Programa de Pós-Graduação em Ciência da Computação da UFRGS no 2º semestre de 1993, foram consideradas adequadas pelos alunos/analistas durante os experimentos.

### 5.3 Conclusões e Pesquisa Futura.

De acordo com o que foi possível observar pela comparação de resultados entre os experimentos 1 e 2, a questão fundamental associada ao benefício potencial da reutilização, é exatamente o grau de similaridade entre os modelos reutilizáveis e os problemas a serem modelados. A correta identificação dos modelos mais similares parece ser fundamental. Os sistemas automatizados de apoio a esta tarefa, podem vir a proporcionar ajuda essencial neste processo. Convém entretanto ressaltar, aliás como já o fez Maiden [MAI 92] quando analisou os aspectos cognitivos associados ao processo de reutilização por analogia, que os sistemas automatizados a exemplo do IRA (Intelligent Reuse Advisor) devem atuar dentro do *paradigma de apoio* [PAL 91 e DIE 92], funcionando como um conselheiro do analista de requisitos.

Entretanto, independentemente de uma possível automatização do processo, alguns aspectos essenciais quanto ao grau de similaridade devem ser analisados:

1. a taxinomia de domínios abstratos proposta por Maiden [MAI 92] baseia-se na funcionalidade dos problemas. Na modelagem com DFD's, a identificação da classe de problemas a qual pertence o problema alvo é um passo decisivo na busca de um problema similar. Mas, como determinar qual, dentre vários possíveis integrantes da mesma classe, é o problema cujo modelo venha a trazer maiores benefícios à reutilização? Que aspectos são de fato importantes na determinação do grau de analogia?

2. no paradigma orientado a objeto, os domínios abstratos de Maiden tem a mesma importância? Ou, alternativamente, o grau de similaridade seria determinado pelo número de objetos/classes compartilhados pelo problema alvo e pelo problema reutilizável?



## BIBLIOGRAFIA

- [ARA 89] ARANGO, G. Domain Analysis - From Art to Engineering Discipline. In: Fifth International Workshop on Software Specification and Design. **Proceedings...** v. 14, n.3, p.152-159, 05-89.
- [CAM 69] CAMPBELL, D. e STANLEY, J. **Experimental and quasi-experimental designs for research**. Chicago, Rand McNally, 1969.
- [CHI 89] CHI, M. T. H., et al. Self-Explanations: How Student Study and Use Examples in Learning to Solve Problems, **Cognitive Science**, n. 13, p. 145-182, 1989.
- [COA 92] COAD, P. e YOURDON, E. **Análise Baseada em Objetos**. Rio de Janeiro: Campus, 1992.
- [DIE 92] DIFENTHAELER, R. **O Desenvolvimento e a Utilização de Sistemas Baseados em Conhecimento Segundo o Paradigma de Apoio**. Dissertação de Mestrado, Instituto de Informática, UFRGS, fevereiro 1992.
- [FRA 94] FRAKES, W. Systematic Software Reuse: a Paradigm Shift. In: THIRD INTERNATIONAL CONFERENCE ON SOFTWARE REUSE, 1994, Rio de Janeiro. **Proceedings...** Los Alamitos, California: IEEE Computer Society Press, 235p., p.2-3, 1994.
- [KAR 89] KARAKOSTAS, V. Requirements for CASE Tools in Early Software Reuse. **Software Engineering Notes**, v.14, n.2, p. 39-41, abril 1989.
- [LEW 91] LEWIS, J.A. et al. An Empirical Study of the Object-Oriented Paradigm and Software Reuse. In: **ACM OOPSLA**, p. 184-196, 1991.
- [MAI 92] MAIDEN, N. A. e SUTCLIFFE, A. G. Exploiting Reusable Specifications through Analogy. **Communication of the ACM**, v.35, n.4, p.55-64 April 1992.
- [MAI 92a] MAIDEN, N. A. Analogical Specification Reuse During Requirements Analysis. **Tese de Doutorado...** : Department of Business Computing, School of Informatics, City University, London, 1992.
- [NOV 88] NOVICK, L. R. Analogical Transfer, Problem Similarity, and Expertise, **Journal of Experimental Psychology: Learning, Memory and Cognition**, 14/3, p.510-520, 1988.
- [PAL 91] PALMER, T. F. e McMENAMIM, S. M. **Análise Essencial de Sistemas**. McGraw-Hill, 1991.
- [PRI 91a] PRIETO-DÍAZ, R. Making Software Reuse Work: An Implementation Model. **ACM SIGSOFT Software Engineering Notes** v.16, n.3, p.61-68, July 1991.
- [REU 89] REUBENSTEIN, H. , B. e WATERS, R. C. The Requirements Apprentice: An Initial Scenario **Software Engineering Notes** v.14, n.5, p.211-8 05-1989.
- [WAT 91] WATERS, R. C. e TAN, Y. M. Toward a Design Apprentice. Supporting Reuse and Evolution in Software Design. **Software Engineering Notes**, v. 16, n. 2, p.33-44, 1991.
- [YOU 90] YOURDON, E. **Análise Estruturada Moderna**. Rio de Janeiro, Campus, 1990.
- [ZIR 93] ZIRBES, S. F. Validação Experimental de Alternativas de Reutilização de Modelos, **Relatório de Pesquisa**, n. 217, PGCC/UFRGS, agosto de 1993.
- [ZIR 95] ZIRBES, S. F. A Reutilização de Modelos de Requisitos de Sistemas por Analogia: Experimentação e Conclusões. **Tese de Doutorado...** Porto Alegre: CPGCC/UFRGS, 1995.