

Evoluindo uma Técnica de Avaliação de Usabilidade através de Estudos *In Vitro* e *In Vivo*

**Marcos Gomes¹, Fábio Santos¹, Davi Viana dos Santos¹,
Guilherme Horta Travassos², Tayana Conte¹**

¹Departamento de Ciência da Computação
Universidade Federal do Amazonas (UFAM) – Manaus, AM - Brasil

²PESC – COPPE/UFRJ, Cx Postal 68.511, CEP 21945-970, Rio de Janeiro, RJ, Brasil
{marcos.sgomes, faerickson, davi.viana}@gmail.com, ght@cos.ufrj.br,
tayana@dcc.ufam.edu.br

Abstract. *This paper presents the development of the usability inspection technique WDP-RT from its conception to its use on industry through experimental studies, in order to guarantee safe transfer of technology from academy to industry. The goals of this paper are: (1) to describe the studies of usability inspection using the technique WDP-RT, in order to encourage the industries in using usability inspections to improve its systems quality; and (2) to spread the knowledge about experimental studies to evaluate and evolve a new Software Engineering technology.*

Resumo. Este artigo apresenta o desenvolvimento da técnica de inspeção de usabilidade WDP-RT desde sua concepção até sua utilização na indústria através de estudos experimentais, com o intuito de garantir a qualidade da tecnologia no processo de transferência da academia para a indústria. Os objetivos deste artigo são: (1) descrever como foram executadas inspeções de usabilidade utilizando a técnica WDP-RT, de forma a incentivar empresas de desenvolvimento a realizar avaliações de usabilidade para melhorar a qualidade de seus produtos; e (2) disseminar o conhecimento sobre o uso de estudos experimentais para avaliar e evoluir uma nova tecnologia de Engenharia de Software.

1. Introdução

Usabilidade é um dos aspectos relacionados à qualidade em uso de interfaces de sistemas, sendo um dos mais importantes critérios de aceitação para aplicações interativas em geral, e em particular para aplicações Web (HITZ *et al.* 2006). Devido à relevância da usabilidade para as aplicações Web, a indústria de desenvolvimento de software está investindo em técnicas e ferramentas para projetos e avaliações que ajudem a melhorar esse atributo de qualidade em suas aplicações Web (MATERA *et al.* 2006). Em termos de avaliação de usabilidade de aplicações Web, várias técnicas específicas têm sido propostas. No entanto, apesar do número de técnicas e ferramentas propostas, muitas organizações de desenvolvimento Web não as estão aplicando (INSFRAN e FERNANDEZ, 2008). Prováveis causas para essa não aplicação são: desconhecimento dessas técnicas e ferramentas; ou ainda limitações de orçamento que impeçam a contratação de especialistas em usabilidade ou a realização de testes de usabilidade. Uma possível causa que precisa ser discutida é a questão de quantas dessas propostas são formuladas através de *princípios científicos* e transferidas de forma segura para a indústria.

Apesar do crescimento da Engenharia de Aplicações Web na última década, o uso do processo científico de experimentação em pesquisas nesta área ainda não está amplamente generalizado. MENDES (2005) discute os resultados de uma revisão na qual foram analisados 173 artigos, dos quais apenas 5% foram considerados adequados do ponto de vista de rigor científico. Vários dos artigos analisados usavam terminologia incorreta, como, por exemplo, “estudo de caso” no lugar de “prova de conceito”. De acordo com MENDES (2005), engenharia é considerada como a aplicação disciplinada de princípios científicos para a resolução de problemas práticos, onde princípios científicos são os resultados da aplicação de um processo científico. O processo científico apóia a construção do conhecimento, que, por sua vez, envolve a utilização de estudos experimentais para testar modelos e hipóteses anteriormente propostos, assegurando que o entendimento atual do campo é correto. Experimentação em Engenharia de Aplicações Web é, portanto, essencial (MENDES *et al.* 2006).

Neste sentido, este artigo apresenta o desenvolvimento de uma técnica de avaliação de usabilidade específica para aplicações Web, chamada WDP-RT (GOMES *et al.* 2009), que foi elaborada como técnica de leitura para poder ser empregada por inspetores com pouco conhecimento sobre usabilidade. O objetivo é que os próprios *stakeholders* de projetos de software possam empregar a técnica na avaliação do software produzido, contribuindo assim para a melhora da sua usabilidade.

A construção e avaliação da técnica WDP-RT se fundamentou em uma abordagem baseada em experimentação, através de uma série de estudos experimentais na academia e na indústria. Além da avaliação experimental, esta abordagem apóia a transferência da nova tecnologia para a indústria de software Web de forma segura, maximizando o benefício de adoção da tecnologia de software em questão.

Os objetivos deste artigo são: (1) descrever como foram executadas inspeções de usabilidade utilizando a técnica WDP-RT e analisar seu custo-eficiência, de forma a incentivar outras empresas de desenvolvimento a realizar avaliações de usabilidade; e (2) disseminar o conhecimento sobre o uso de estudos experimentais para avaliar e evoluir uma nova tecnologia de Engenharia de Software.

O restante deste artigo está estruturado da seguinte forma. A seção 2 apresenta a base para a proposta inicial da WDP-RT. A seção 3 apresenta os estudos experimentais realizados e como eles apoiaram a evolução da WDP-RT. A seção 4 pondera sobre os benefícios trazidos para a indústria. A seção 5 discorre sobre as lições aprendidas nos estudos realizados. Por fim, a seção 6 conclui o artigo.

2. Proposta inicial da técnica WDP-RT (Web Design Perspectives-based Inspection – Reading Technique)

Os métodos de inspeção se baseiam em três tipos de técnicas de inspeção: *ad hoc*, *checklist* e técnica de leitura. Na inspeção *ad hoc*, o revisor não utiliza nenhuma técnica sistemática. Numa técnica de inspeção baseada em *checklist*, os inspetores recebem uma lista de verificação (*checklist*) que os auxilia a encontrar os defeitos. De acordo com TRAVASSOS *et al.* (2002), uma técnica de leitura é um tipo específico de técnica de inspeção que contém uma série de passos para a análise individual de um produto de software de forma a alcançar a compreensão necessária para uma tarefa específica

A técnica WDP-RT é originada de outra técnica de inspeção, a técnica WDP (CONTE *et al.* 2007), também desenvolvida a partir de uma metodologia baseada em

evidências (SHULL *et al.* 2001; MAFRA *et al.* 2006) A técnica WDP utiliza as heurísticas de NIELSEN (1994) como base, direcionando a avaliação de usabilidade através de perspectivas específicas para a representação de aplicações Web: Apresentação, Conceituação e Navegação.

Por ser uma técnica de inspeção baseada em *checklist*, a WDP se adéqua facilmente ao uso por inspetores com alguma experiência em inspeções de usabilidade, onde eles podem executar a inspeção da forma mais conveniente, e contam com a ajuda do *checklist* para verificação de características pontuais na aplicação Web. Entretanto, como apresentado no estudo de observação relatado em (CONTE *et al.* 2009b), isto ainda não é suficiente para ajudar inspetores novatos na execução desta atividade. Inspetores novatos precisam de direcionamento, um procedimento a ser seguido para executar a inspeção. A WDP fornece as diretrizes para a avaliação, porém não informa aos inspetores em que ordem estas diretrizes ou mesmo as perspectivas de projeto Web devem ser aplicadas.

A WDP-RT é uma extensão da WDP para uma técnica de leitura. Ela se baseia em um conjunto de instruções que devem ser executadas para a verificação da usabilidade da aplicação. Com o propósito de aumentar a cobertura de avaliação da WDP, as instruções da WDP-RT foram definidas realizando-se uma análise de equivalência de características (*feature analysis*) de três conjuntos de características de usabilidade: conjunto de itens a serem verificados propostos pela WDP, com base nas heurísticas de (NIELSEN, 1994), os “requisitos não funcionais de usabilidade” (FERREIRA E LEITE, 2003), e o conjunto de “características funcionais de usabilidade” (JURISTO *et al.* 2007).

As instruções da WDP-RT estão agrupadas de acordo com as três perspectivas de projeto Web, sendo executadas primeiramente as instruções para a verificação da usabilidade em relação à perspectiva Apresentação, seguida das instruções referentes à perspectiva Conceituação, e por fim, à perspectiva Navegação. A Figura 1 apresenta uma visão da primeira versão da WDP-RT (WDP-RT v1). O texto completo da WDP-RT v1 está disponível em (GOMES e CONTE, 2009a).

3. Para cada página que possuir entrada de dados (preenchimento de caixas de texto, itens de seleção, etc.):
 - a. Observe se os dados obrigatórios na entrada de dados estão claramente definidos (por exemplo, marcados com um *).
 - b. Observe os campos a serem preenchidos. Existe informação sobre o formato correto para entrada de dados específica? Por exemplo, dd/mm/aaaa para campos do tipo data. O sistema facilita a entrada de dados simples ou estruturalmente complexos?

Figura 1 - Extrato da WDP-RT v1

3. Execução de Estudos Experimentais para Evolução da WDP-RT

Estudos experimentais são os “blocos de construção do conhecimento” necessários para construir evidências e determinar quais as melhores situações para empregar determinada tecnologia. TRAVASSOS E BARROS (2003) apresentam uma classificação para os estudos experimentais, considerando as diferenças de acordo com o nível de controle existente sobre o ambiente onde são executados, sendo classificados em: *in vivo*, *in vitro*, *in virtuo*, e *in silico*.

Estudos *in vitro* são realizados em um ambiente controlado, com a participação de pessoas que atuam como representantes da população de interesse. Estudos *in vivo* envolvem pessoas em seu próprio ambiente de trabalho em condições realistas. Estudos *in virtuo* e *in silico* possuem como característica a exploração de simulações do mundo real, conduzidas em ambientes virtuais compostos por modelos computacionais.

Para o desenvolvimento e evolução da WDP-RT, foram realizados dois estudos *in vitro* e um estudo *in vivo*, que serão apresentados a seguir.

3.1. Primeiro estudo de viabilidade

Segundo a metodologia para transferência de tecnologias para a indústria proposta por SHULL *et al.* (2001), o primeiro estudo que deve ser realizado para avaliar uma nova tecnologia é um estudo de viabilidade, que visa verificar se a nova tecnologia é viável e se o tempo empregado é bem utilizado. Para responder estas perguntas, foi realizada uma inspeção de usabilidade comparando a viabilidade da WDP-RT com os resultados apresentados pela WDP. Este estudo de viabilidade, realizado em Junho de 2009, pode ser caracterizado como um estudo *in vitro*, pois foi executado em um ambiente controlado e teve como participantes, em sua maioria, alunos no lugar de inspetores. Este primeiro estudo foi discutido detalhadamente em (GOMES *et al.* 2009) e será apresentado um resumo dos resultados do mesmo.

Participantes: Dezoito alunos da disciplina de Engenharia de Software, do curso de graduação em Ciência da Computação da Universidade Federal do Amazonas, além de duas desenvolvedoras de software de uma instituição de pesquisa. Os participantes foram divididos em dois grupos de dez inspetores (equipes WDP e WDP-RT) com níveis de experiência balanceados de acordo com as respostas dos formulários de caracterização preenchidos por eles.

Objeto de estudo: A aplicação inspecionada foi o portal da Sociedade Brasileira de Computação (SBC)¹, para o qual foi definido um roteiro com oito atividades a serem desempenhadas no portal, tais como informar-se sobre o POSCOMP, associar-se a SBC e informar-se sobre o Simpósio Brasileiro de Engenharia de Software de 2009.

Procedimento: Cada grupo recebeu um treinamento de uma hora que incluía conceitos sobre usabilidade e a técnica específica que o grupo deveria utilizar para realizar a inspeção. Cada inspetor pôde realizar a detecção individual no horário que lhe fosse mais conveniente, respeitando-se apenas o prazo de dois dias.

Coleção de dados: Dezesseis inspetores (nove do grupo WDP e sete do grupo WDP-RT) enviaram as suas planilhas de anotação de discrepâncias e o questionário de avaliação da técnica. Algumas planilhas apresentaram problemas (alguns inspetores não informaram o tempo de inspeção ou não seguiram o roteiro) e foram descartadas. Desta forma, foram analisadas seis planilhas do grupo WDP e seis do grupo WDP-RT.

As listas de discrepâncias individuais foram integradas a uma única lista e uma equipe de três pesquisadores realizou uma reunião para decidir quais dessas discrepâncias eram únicas e quais eram duplicatas (discrepâncias equivalentes apontadas por mais de um inspetor). A reunião de discriminação contou com a participação dos três pesquisadores que elaboraram a lista única e de um diretor da SBC. Cada discrepância relatada foi avaliada e após a discussão da equipe, o diretor da SBC classificava a discrepância em defeito ou falso-positivo.

1 – www.sbc.org.br

Resultados obtidos: Neste estudo, foram analisadas a eficácia e a eficiência da WDP-RT em comparação com a WDP. A eficácia é definida como a razão entre o número de defeitos detectados pela técnica e o número total de defeitos conhecidos. A eficiência é definida como a razão entre o número total de defeitos (incluindo as duplicatas) e o tempo gasto na inspeção. A análise estatística apontou que a WDP-RT foi mais eficaz que a WDP, além de apresentar eficiência semelhante à WDP. A Tabela 1 sumariza os resultados.

Tabela 1: Resultados do 1º estudo de viabilidade

Técnica	Total de Defeitos	Eficácia	Defeitos por Inspetor	Média de tempo (h)	Defeitos / hora	DP ² (Defeitos / Inspetor)	%DP ² (Defeitos / Inspetor)
WDP	55	63,01	9,16	1,23	7,44	2,31	25,27%
WDP-RT	127	90,41	21,16	2,86	7,39	7,83	37%

Os inspetores também responderam um questionário para verificar sua opinião em relação à técnica. A análise destes questionários está resumida a seguir:

- Os inspetores consideram a WDP-RT útil para inspeções de usabilidade e de fácil utilização. Todos consideraram as perspectivas de projeto Web e as instruções da WDP-RT fáceis de compreender.
- Um dos inspetores, após aplicar a técnica linearmente para as quatro primeiras atividades, assimilou as diretrizes, passando a primeiro procurar pelos defeitos e depois associá-los às instruções.
- Dois inspetores relataram que algumas instruções eram semelhantes.
- Metade dos inspetores ressaltou o alto tempo de inspeção.

3.2.Evolução da WDP-RT

A análise quantitativa já apontava o elevado tempo de inspeção dos inspetores que utilizaram a WDP-RT, e após a análise qualitativa, foi feita a primeira revisão da técnica. Esta revisão teve como principais objetivos procurar por instruções ambíguas e estudar uma forma de tornar a inspeção com WDP-RT mais rápida.

Após análise detalhada da técnica, verificou-se que as instruções das perspectivas de Apresentação e Conceituação eram bem semelhantes, porém o foco da instrução era direcionado para a Apresentação ou para a Conceituação. Um exemplo, citado por um dos inspetores, são as instruções 2g (Apresentação) e 8c (Conceituação):

- 2g - Observe se a página disponibiliza opções de ajuda e documentação ao usuário e se ela está claramente visível para ele. Acesse a opção de ajuda e verifique se as informações de ajuda e documentação são apresentadas de forma que auxiliem o usuário a atingir seus objetivos na página.
- 8c - Observe se a página disponibiliza opções de ajuda e documentação ao usuário. Esta ajuda é compreensível pelo usuário? Ajuda o usuário na realização de tarefas complexas ou no entendimento de conceitos não triviais?

Mesmo que as instruções apresentem focos diferentes, o inspetor teria que executar a mesma tarefa (verificar opções de ajuda e documentação) duas vezes para realizar a inspeção, sendo este um dos possíveis fatores que tornariam a inspeção com WDP-RT mais demorada.

Para diminuir o número de instruções semelhantes e tentar reduzir o tempo de inspeção, decidiu-se pela junção das perspectivas de Apresentação e Conceituação. As instruções semelhantes das duas perspectivas foram agrupadas, de forma que o inspetor possa verificar em paralelo a usabilidade em relação à Apresentação e à Conceituação. A seguir é apresentado o exemplo da nova instrução 2g, que representa a junção das instruções 2g e 8c da WDP-RT v1:

- Observe se a página disponibiliza opções de ajuda e documentação ao usuário e se ela está claramente visível para ele. Acesse a opção de ajuda e verifique se as informações de ajuda e documentação são apresentadas de forma que auxiliem o usuário a atingir seus objetivos na página. Esta ajuda é compreensível pelo usuário? Ajuda o usuário na realização de tarefas complexas ou no entendimento de conceitos não triviais?

Desta forma, a inspeção com WDP-RT passa a ter duas etapas: na primeira parte, o inspetor verifica a usabilidade em relação às perspectivas de Apresentação e Conceituação, e na segunda parte verifica a usabilidade em relação à perspectiva de Navegação. Foram revisadas as definições sobre o objetivo de cada uma das etapas de inspeção, para deixar claro para o inspetor o objetivo de inspeção em paralelo das duas perspectivas. O texto da instrução 4 (referente a opções de personalização da aplicação Web) também foi modificado, para deixar a cargo do inspetor decidir se deve ou não aplicá-la.

Estas modificações resultaram na segunda versão da WDP-RT (WDP-RT v2). O texto completo da WDP-RT v2 está disponível em (GOMES e CONTE, 2009b).

3.3.Segundo estudo de viabilidade

No primeiro estudo de viabilidade, muitos participantes apresentavam bons conhecimentos em usabilidade, inspeções e desenvolvimento de aplicações Web. Para investigar se a WDP-RT realmente ajuda inspetores com pouco conhecimento em inspeções de usabilidade, foi realizado um segundo estudo de viabilidade. O estudo constituiu-se de uma inspeção de usabilidade utilizando-se as técnicas WDP e WDP-RT v2, na qual foram avaliadas a eficácia e eficiência de ambas as técnicas. Esta inspeção ocorreu em Outubro de 2009, no âmbito da disciplina de Modelagem e Projeto de Sistemas, do curso de graduação em Ciência da Computação da Universidade Federal do Amazonas.

Doze alunos voluntários matriculados na disciplina participaram do estudo. Todos os participantes assinaram um Termo de Consentimento Livre e Esclarecido e preencheram um formulário de caracterização que continha perguntas que deveriam ser respondidas em uma escala de 1 (nenhuma experiência prática) a 5 (experiência em vários projetos industriais) sobre o conhecimento dos participantes sobre usabilidade, avaliações e inspeções de software e em desenvolvimento de software Web. Os níveis de experiência para estes 12 participantes variaram apenas entre “Nenhuma” e “Baixa Experiência, conforme apresentado na Tabela 2. Os participantes foram divididos em dois grupos de seis inspetores (equipes WDP e WDP-RT) com níveis de experiência balanceados de acordo com as respostas dos formulários de caracterização.

A aplicação inspecionada foi a mesma aplicação inspecionada no primeiro estudo de viabilidade, o portal da Sociedade Brasileira de Computação (SBC). Entretanto, a quantidade de atividades a serem desempenhadas pelos inspetores foi

restringida às três atividades que apresentaram o maior número de defeitos no estudo anterior: informar-se sobre o POSCOMP, associar-se a SBC e cadastrar-se no portal.

Cada grupo de participantes recebeu um treinamento com duração de uma hora, que incluía conceitos sobre usabilidade e a técnica específica que o grupo deveria utilizar para realizar a inspeção. Cada participante recebeu o roteiro com as atividades a serem inspecionadas, a técnica de inspeção estudada no treinamento impressa, uma planilha para a anotação das discrepâncias encontradas, além da própria apresentação do treinamento. Além disso, eles também receberam um questionário de avaliação da técnica estudada. Cada inspetor pôde realizar a detecção individual no horário que lhe fosse mais conveniente, respeitando-se apenas o prazo de uma semana.

3.3.1. Execução da Inspeção

No prazo estipulado, os doze inspetores enviaram as suas planilhas de anotação de discrepâncias e o questionário de avaliação da técnica. As planilhas de discrepâncias individuais foram então integradas a uma única lista, retirando-se a referência do inspetor e da técnica utilizada por ele. Para facilitar a discriminação das discrepâncias, decidiu-se utilizar a lista de discrepâncias do primeiro estudo de viabilidade como “oráculo”.

Dois pesquisadores em Engenharia de Software, que não estão relacionados com a pesquisa das técnicas envolvidas neste estudo, realizaram uma reunião para decidir quais dessas discrepâncias eram únicas e quais eram duplicatas. Os pesquisadores também tentavam associar as novas discrepâncias às discrepâncias do “oráculo” (os pesquisadores receberam uma cópia do “oráculo” sem a classificação da discrepância em defeito ou falso-positivo). As discrepâncias que não puderam ser associadas foram analisadas e, após discussão, classificadas em defeito ou falso-positivo.

3.3.2. Resultados obtidos

Assim como no primeiro estudo de viabilidade, foi medido o número de defeitos encontrados por cada inspetor na avaliação do portal da SBC.

Tabela 2: Resultados por inspetor

Técnica/ Inspetor	Experiência/ Usabilidade	Experiência/ Inspeções	Experiência/ Desenvolvimento	# Disc. ³	# FP ⁴	# Def. ⁵	Tempo	Def./ Hora	Total Defeitos	
WDP	01	Nenhum	Nenhum	Nenhum	30	3	27	106	15,28	80
	02	Nenhum	Nenhum	Baixo	14	0	14	90	9,33	
	03	Nenhum	Nenhum	Baixo	16	2	14	150	5,60	
	04	Nenhum	Nenhum	Nenhum	11	2	9	132	4,09	
	05	Baixo	Nenhum	Nenhum	8	2	6	75	4,80	
	06	Baixo	Nenhum	Baixo	11	1	10	78	7,69	
WDP -RT	11	Baixo	Nenhum	Nenhum	41	4	37	202	10,99	107
	12	Nenhum	Nenhum	Nenhum	24	1	23	240	5,75	
	13	Nenhum	Nenhum	Nenhum	12	3	9	70	7,71	
	14	Baixo	Baixo	Baixo	15	2	13	65	12,00	
	15	Baixo	Nenhum	Nenhum	20	1	19	157	7,26	
	16	Baixo	Nenhum	Baixo	10	4	6	98	3,67	

O indicador eficácia é definido como a razão entre o número de defeitos detectados pela técnica e o número total de defeitos conhecidos. No primeiro estudo de viabilidade, foram encontrados 73 defeitos, dos quais 49 estão relacionados às atividades do roteiro deste segundo estudo. Neste segundo estudo, foram encontrados 86

3 – número de discrepâncias, 4 – número de falso-positivos, 5 – número de defeitos

defeitos, dos quais 47 foram relatados apenas neste segundo experimento. Desta forma, considerando-se todos os defeitos relatados nos dois estudos de viabilidade, foi encontrado um total de 96 defeitos. A Tabela 3 compara a eficácia das técnicas.

Tabela 3: Comparativo de eficácia

Técnica	Defeitos	Total de defeitos conhecidos	Eficácia
WDP	50	96	52,08
WDP-RT	53		55,20

Para analisar o indicador eficiência, é necessário considerar todos os defeitos (incluindo-se também as duplicatas) encontrados. A eficiência então é definida como a razão entre o número total de defeitos e o tempo gasto na inspeção. A Tabela 4 mostra o comparativo de eficiência para ambas as técnicas.

Tabela 4: Comparativo de eficiência

Técnica	Total de Defeitos	Defeitos por Inspetor	Média de tempo (hora)	Média de Defeitos / hora	DP ⁶ (Defeitos por Inspetor)	%DP ⁶ (Defeitos por Inspetor)
WDP	80	13,33	1,75	7,60	7,36	55,24
WDP-RT	107	17,83	2,31	7,71	11,28	63,28

Analisando a tabela 2, verifica-se que os inspetores que utilizaram a WDP-RT encontraram um número maior de defeitos (107 x 80), assim como apresentado no primeiro estudo de viabilidade. Entretanto, tanto a eficácia quanto a eficiência foram semelhantes às da WDP, conforme verificado nas tabelas 3 e 4.

Em relação ao tempo de inspeção, observando as tabelas 6 e 8, verifica-se que a inspeção com WDP-RT demorou, em média, 31,85% a mais em comparação com o tempo médio de inspeção com WDP. Entretanto, no primeiro estudo de viabilidade, os inspetores que utilizaram WDP-RT levaram, em média, mais que o dobro de tempo (132%) que a inspeção com WDP (Tabela 1). Mesmo sendo uma amostra pequena, isto é um indício de que a WDP-RT v2 possibilita que a inspeção seja mais ágil, em comparação com a WDP-RT v1.

Para coletar a opinião dos inspetores em relação à facilidade de uso e a utilidade da técnica WDP-RT, e a facilidade de compreensão das instruções e das perspectivas, os inspetores responderam um questionário ao término da inspeção. A análise destes questionários está sumarizada a seguir:

- Os inspetores consideram a WDP-RT útil para inspeções de usabilidade. Ainda consideraram as perspectivas de projeto Web e as instruções da WDP-RT fáceis de compreender.
- Dois inspetores relataram que algumas instruções eram repetitivas.
- Em relação à facilidade de uso, dois inspetores relataram dificuldades em realizar a inspeção, um por achar algumas instruções de difícil entendimento, outro por achar a técnica extensa.
- Três inspetores afirmaram ter seguido a técnica linearmente. Outros dois inspetores leram a técnica, buscando decorá-la. Após isso, realizaram uma

inspeção *ad hoc* e tentavam associar os defeitos encontrados às instruções da WDP-RT.

3.3.3. Ameaças à Validade dos Estudos *In Vitro*

Em todos os estudos experimentais existem ameaças que podem afetar a validade dos resultados. As ameaças relacionadas aos estudos *in vitro* são apresentadas a seguir:

Validade Interna: foram consideradas três ameaças: (1) efeitos de treinamento, (2) classificação de experiência e (3) medição de tempo. Em relação à primeira ameaça, poderia haver um efeito causado pelo treinamento, caso o treinamento da técnica WDP tivesse qualidade inferior ao treinamento da WDP-RT. Este risco foi controlado preparando treinamentos equivalentes com os mesmos exemplos de problemas detectados aplicando WDP ou WDP-RT. Em relação à classificação de experiência dos participantes, ela foi uma auto-classificação, com base em número e tipo de experiências anteriores (experiência acadêmica e na indústria). Por fim, sobre a medição do tempo, foi solicitado aos participantes que eles fossem precisos nas suas medições, porém não há garantia que o tempo relatado foi realmente medido cuidadosamente.

Validade Externa: três questões foram consideradas: (1) provavelmente estudantes não são bons substitutos para inspetores da indústria; (2) normalmente ambientes acadêmicos não simulam totalmente as condições existentes em ambientes industriais; e (3) validade do portal SBC como representante de aplicações Web. Sobre a questão (1), segundo CARVER *et al.* (2003), mesmo estudantes que não possuem experiência em aplicações na indústria podem apresentar habilidades similares a inspetores menos experientes. Em relação à questão (2), o portal SBC é uma aplicação Web real. Sobre a questão (3), não é possível afirmar que o portal SBC represente todo tipo de aplicação Web, uma vez que são várias as categorias de aplicações (KAPPEL *et al.* 2006).

Validade de Conclusão: o maior problema é o tamanho da amostra, com um número pequeno de *data points*, não ideal do ponto de vista estatístico. Devido a este fato, há limitação nos resultados, sendo estes considerados não conclusivos, e sim indícios.

Validade de Constructo: referente a este tipo de ameaça de validade, considerou-se a definição dos indicadores. Os indicadores adotados neste estudo - Eficiência e Eficácia - são comumente utilizados em estudos que investigam técnicas de detecção de defeitos e estes indicadores foram medidos utilizando a mesma abordagem proposta por DINGER e KOLB (2006) e aplicada em (CONTE *et al.* 2007).

3.4. Estudo de caso na indústria

Com a realização dos estudos *in-vitro* conseguiu-se indícios da viabilidade da técnica WDP-RT. Desta forma, pode-se planejar um novo estudo, desta vez um estudo *in vivo*, para avaliar a adequação da WDP-RT ao ambiente industrial. Decidiu-se então pela realização de um estudo de caso, visto que estes permitem a análise de um processo particular no contexto de um ciclo de vida de software (SHULL *et al.* 2001). Com este tipo de estudo, espera-se responder a seguinte questão de pesquisa: “A técnica é adequada no contexto de um ciclo de vida?”. O estudo de caso foi realizado nos meses de Dezembro de 2009 e Janeiro de 2010 e contou com a colaboração da empresa FabriQ Informática Ltda⁷. O objetivo deste estudo, de acordo com o paradigma GQM (BASILI e ROMBACH, 1988), é apresentado na Figura 2.

7 - www.fabriq.com.br

Analisar	Técnica WDP-RT
Com o propósito de	avaliar
Em relação a	adequação da utilização da WDP-RT em um ciclo de vida real
Do ponto de vista	Desenvolvedores de Software, analistas de suporte e Avaliadores de Requisitos de Aplicações Web com baixa experiência em inspeções de usabilidade
No contexto de	uma avaliação de usabilidade de módulos de um sistema organizacional Web por seus desenvolvedores e avaliadores de requisitos.

Figura 2 – Objetivo do estudo de caso na Indústria

A FabriQ é uma empresa de desenvolvimento de pequeno porte inserida dentro do Programa AMAZONSOFT e residente no Centro de Incubação e Desenvolvimento Empresarial - CIDE. É uma empresa pioneira no mercado de Tecnologia de Informação na Região Norte, especialmente no desenvolvimento de softwares para Qualidade. Atualmente está em fase de implantação da ISO 9000 e do MPS-BR nível G.

O objeto deste estudo foi o Módulo de Controle de Documentos da aplicação DOMMA ISO. O DOMMA ISO é um software de gerenciamento e controle de fluxo de documentos, além do tratamento de não-conformidades e planejamento de auditorias e planos de ação. Este software atende os requisitos essenciais da ISO 9000. O módulo inspecionado é composto de 14 casos de uso, nos quais atuam 5 perfis de usuário. Para facilitar a inspeção, os casos de uso foram divididos em dois roteiros de atividades (A e B), onde o roteiro A continha os casos de uso referentes a três perfis e o roteiro B continha os casos de uso referentes aos outros dois perfis.

Os participantes do estudo foram oito profissionais de informática (cinco analistas de sistemas e três analistas de suporte). Cada um dos participantes assinou um Termo de Consentimento Livre e Esclarecido e preencheu um formulário de caracterização contendo perguntas sobre o seu conhecimento em usabilidade, avaliações e inspeções de software e em desenvolvimento de software Web. O formulário de caracterização também continha questões referentes à relação do profissional com o módulo da aplicação a ser inspecionado (participação na análise, programação ou teste dos casos de uso do módulo). Os participantes foram divididos em dois grupos de quatro inspetores (equipes A e B) com níveis de experiência balanceados de acordo com as respostas dos formulários de caracterização, conforme apresentado na Tabela 5.

Os participantes receberam um treinamento com duração de uma hora e quinze minutos, que continha conceitos sobre usabilidade e a técnica WDP-RT. Ao término do treinamento, cada participante recebeu o roteiro com as atividades a serem inspecionadas (de acordo com o seu grupo), a técnica WDP-RT impressa, uma planilha para a anotação das discrepâncias encontradas, além da própria apresentação do treinamento. Além disso, eles também receberam um questionário de avaliação da técnica estudada. A detecção de defeitos foi realizada individualmente pelos inspetores, que tiveram um prazo de uma semana para sua realização.

3.4.1. Execução da inspeção

No prazo estipulado, sete inspetores (quatro do grupo A e três do grupo B) enviaram as suas planilhas de anotação de discrepâncias e o questionário de avaliação da técnica. Um dos inspetores não pôde realizar a inspeção por precisar se dedicar a outras atividades durante este período. As sete planilhas enviadas foram analisadas.

Por falta de tempo dos colaboradores da organização, os pesquisadores envolvidos no estudo realizaram a coleção das discrepâncias relatadas. Para cada roteiro

de atividades foi gerada uma lista única, contendo todas as discrepâncias relatadas nas planilhas de discrepâncias individuais dos inspetores do roteiro. Os pesquisadores retiraram a referência do inspetor e analisaram quais dessas discrepâncias eram únicas e quais eram duplicatas (discrepâncias equivalentes apontadas por mais de um inspetor).

Foi realizada uma reunião de discriminação para cada roteiro. Cada reunião contou com a participação dos inspetores que executaram a inspeção do roteiro e de dois responsáveis pelo sistema. Os pesquisadores também participaram da reunião como mediadores. Nesta reunião, as interações avaliadas foram re-executadas, sendo assim possível verificar *in-loco* cada discrepância relatada. Após a discussão da discrepância entre os inspetores e responsáveis, esta era classificada em defeito ou falso-positivo.

3.4.2. Resultados obtidos

O primeiro objetivo deste experimento foi verificar a adequação da WDP-RT ao ambiente industrial. Para isto, foi observado o número de defeitos apontados por cada um dos inspetores. Três inspetores já apresentavam algum conhecimento prévio em usabilidade e inspeções, sendo que dois deles já haviam participado de inspeções de usabilidade. Em relação a este objetivo, o resultado foi positivo, como pode ser observado na Tabela 5.

Tabela 5: Resultados por inspetor

Roteiro / Inspetor	Experiência Usabilidade	Experiência Inspeções	Experiência Desenvolvimento	# Disc. ⁸	# FP ⁹	# Def. ¹⁰	Tempo	Def./ Hora	Total Defeitos	
A	10	Nenhum	Nenhum	Baixo	5	1	4	35	6,85	28
	11	Nenhum	Nenhum	Baixo	4	0	4	70	3,42	
	12	Médio	Médio	Alto	12	1	11	60	11,00	
	13	Baixo	Baixo	Médio	12	3	9	135	4,00	
B	14	Nenhum	Nenhum	Alto	22	1	21	150	8,40	56
	15	Alto	Médio	Baixo	27	6	21	120	10,50	
	17	Nenhum	Nenhum	Baixo	15	1	14	75	11,20	

O indicador eficácia não pôde ser calculado, visto que o número total de defeitos do módulo inspecionado não era conhecido. Quanto ao indicador eficiência, este foi calculado como a razão entre o número de defeitos encontrados por tempo de inspeção. Foram encontrados 84 defeitos durante a inspeção. Em média, os inspetores gastaram 1 hora e 32 minutos na detecção. Assim, a *Eficiência na etapa de Detecção* foi de 7,81 defeitos / hora por inspetor.

Outro indicador calculado neste estudo foi o *Esforço nas etapas de Detecção e Discriminação*. Para calcular este indicador, é necessário considerar o tempo gasto na atividade de discriminação. Em relação ao tempo gasto na etapa de discriminação, foram realizadas duas reuniões, uma para cada roteiro. A primeira reunião teve duração de 1 hora e a segunda reunião teve duração de 1 hora e 40 minutos. Neste estudo de caso, o custo da inspeção se mostrou baixo, pois o esforço médio de um inspetor, somando o esforço na atividade de detecção (1 hora e 32 minutos) e discriminação (1 hora e 20 minutos) foi de 2 horas e 52 minutos.

O terceiro indicador calculado foi o *Grau de Apreensibilidade*. Este indicador denota o quão fácil é aprender um método (BOLCHINI e GARZOTTO, 2007). De forma análoga à (BOLCHINI e GARZOTTO, 2007), este indicador foi verificado através do fator tempo gasto no treinamento da técnica. O tempo de treinamento da WDP-RT foi de apenas 1 hora e 15 minutos.

8 – número de discrepâncias, 9 – número de falso-positivos, 10 – número de defeitos

Sobre os benefícios obtidos para o software e a organização (FabriQ), além da melhora da usabilidade do módulo inspecionado, a inspeção proporcionou uma melhoria do padrão de interação adotado no sistema, o que acarreta automaticamente benefícios em relação à usabilidade de todos os módulos do DOMMA ISO. Adicionalmente, como resultado indireto, espera-se o aumento do conhecimento dos inspetores sobre usabilidade.

3.4.3. Análise Qualitativa

Após as reuniões de discriminação, os inspetores foram entrevistados com o objetivo de coletar suas opiniões em relação à WDP-RT. Estas entrevistas foram individuais e realizadas pelos dois pesquisadores envolvidos no estudo. Os principais pontos levantados nas entrevistas estão listados a seguir:

- Apenas um inspetor seguiu a técnica linearmente. Os inspetores normalmente liam a técnica e tentavam memorizá-la. Após isto, eles buscavam pelos problemas de usabilidade e tentavam associá-los às instruções da WDP-RT.
- Quatro inspetores consideraram a técnica de fácil utilização, orientando a inspeção. Outros três inspetores relataram dificuldades em aplicar a técnica. Entretanto, quando questionados sobre a dificuldade, eles apontaram a associação dos defeitos encontrados com os pares HxP da WDP, que eles haviam visto no treinamento. Notou-se então que esse fato havia causado um viés no aprendizado da WDP-RT.
- Em relação à inspeção em paralelo das perspectivas de Apresentação e Conceituação, quatro inspetores consideraram que isto facilita a inspeção e a torna mais rápida. Porém, outros três inspetores apontaram que isto pode causar dificuldades de entendimento.
- Todos os inspetores consideraram que a WDP-RT facilita a inspeção. Ela ajuda a melhorar os conhecimentos sobre usabilidade, e eles presumem que isto os ajudará no desenvolvimento de sistemas com melhor usabilidade. Todos também consideraram que a WDP-RT poderia ser utilizada para a realização de inspeções nos produtos desenvolvidos pela empresa.

4. Benefícios para a Indústria na Adoção de Avaliações de Usabilidade

No cenário de desenvolvimento de aplicações de software convencionais, é possível constatar aumentos substanciais de qualidade e redução de defeitos devido à adoção de inspeções e outras técnicas de revisão (TRAVASSOS *et al.* 1999).

A usabilidade aumenta a satisfação e produtividade do cliente, leva à confiança e fidelidade do usuário, conduzindo à redução de custos e aumento de lucros. Assume importância na imagem corporativa da empresa, afetando o valor da marca e quota de mercado (MARCUS, 2002). Muitos sites afastam quase metade dos atuais clientes, por não disponibilizarem facilmente a informação que eles precisam (NIELSEN, 1997).

As inspeções melhoram a produtividade uma vez que os defeitos são encontrados quando são mais fáceis e mais baratos para corrigir. No estudo *in vivo* relatado, o custo da inspeção se mostrou baixo, não sendo necessária a contratação de um especialista em usabilidade e nem utilização de ambiente especial. Foi possível realizar a inspeção de usabilidade com profissionais de informática da própria empresa, que só precisaram de um treinamento de 1 hora e 15 minutos para realizar a inspeção. O

tempo dedicado à detecção e discriminação foi em média de 2 horas e 52 minutos. Desta forma, a empresa foi capaz de realizar a inspeção investindo somente 4 horas e 07 minutos dos analistas de sistemas e suporte disponibilizados para esta atividade.

Além dos benefícios já citados, a utilização da WDP-RT proporciona direcionamento para a execução da inspeção de usabilidade, tornando a inspeção mais fácil para os inspetores com pouco conhecimento em usabilidade. Segundo os inspetores que participaram do estudo de caso na indústria, a técnica ajudou-os a melhorar os seus conhecimentos sobre usabilidade, o que se espera contribuir para que os próximos aplicativos sejam desenvolvidos com melhor usabilidade. Ainda segundo os inspetores, o aumento da usabilidade aumentará a credibilidade dos produtos e reduzirá as chamadas ao serviço de atendimento ao cliente, além da redução de correções dos aplicativos.

5. Lições Aprendidas nos Estudos Realizados

Um dos objetivos deste artigo é compartilhar o conhecimento sobre o uso de estudos experimentais para avaliar e evoluir novas tecnologias para Engenharia de Software. Os estudos realizados possibilitaram o aprendizado de lições específicas sobre a condução de estudos experimentais sobre inspeções e técnicas de leitura. A seguir compartilhamos os principais aprendizados com o intuito de apoiar futuros estudos experimentais:

- Classificação dos inspetores: Em todos os experimentos *in vitro* apresentados neste artigo, a classificação dos inspetores baseou-se em um formulário de caracterização que continha questões sobre o conhecimento deles em relação à usabilidade, inspeções e desenvolvimento Web. Entretanto, verificou-se que mesmo inspetores que possuíam caracterizações de experiência semelhantes apresentaram resultados diferentes em termos de custo-eficiência, conforme pode ser verificado na tabela 2 e em (GOMES *et al.* 2009). Embora o baixo número de participantes destes estudos não permita gerar análises conclusivas, estes resultados podem indicar a necessidade de estudar o impacto de outras variáveis, além da experiência, em relação à classificação do inspetor. Uma possível questão de pesquisa é averiguar se há relacionamento entre habilidades e perfis comportamentais dos profissionais que indique maior eficiência no papel de inspetor, a exemplo da pesquisa sobre o papel de SQA em relação à Teoria de Papéis em Time de Belbin (MEIRA e SILVA, 2009);
- Utilização de técnicas de leitura: Para avaliar uma técnica de leitura, é essencial que aquele que a utilize siga as instruções passo a passo, pelo menos no início do processo. Nos estudos realizados, verificou-se que muitos inspetores que utilizaram a WDP-RT não a seguiam corretamente, o que acaba comprometendo a avaliação. Para poder avaliar esta situação, acreditamos que o ideal seja realizar estudos de observação, onde os próprios pesquisadores acompanhem a aplicação da técnica;

6. Conclusões e trabalhos futuros

Este artigo apresentou uma técnica de leitura para inspeção de usabilidade de aplicações Web, a WDP-RT, e os estudos experimentais realizados para a transferência segura da tecnologia da academia para a Indústria. Os estudos realizados apontam que a WDP-RT facilita a inspeção de usabilidade e, embora os resultados de um único estudo de caso não possam ser generalizados para outros contextos, o resultado qualitativo e

quantitativo do estudo *in vivo* é um indicador da viabilidade de utilização pelos próprios membros da equipe de um projeto como inspetores em uma avaliação de usabilidade. Espera-se com estes resultados incentivar a indústria de software Web a realizar avaliações de usabilidade com maior frequência.

De acordo com TRAVASSOS (2008), a experimentação pode apoiar o desenvolvimento, evolução e melhoria dos processos e tecnologias de software. Ao apresentar os estudos experimentais realizados para evolução da WDP-RT, almeja-se contribuir com o compartilhamento do conhecimento científico, objetivando melhor interação entre a academia e a indústria de software.

Como trabalhos futuros, temos a evolução da WDP-RT para uma nova versão, considerando os resultados quantitativos e qualitativos obtidos nos estudos realizados, a fim de tornar a inspeção mais ágil. Para avaliar a utilização da técnica pelos inspetores, será realizado um estudo de observação, já com a nova versão da WDP-RT. Este estudo visa coletar dados de como os inspetores aplicam a técnica. Também serão realizados novos estudos *in vivo*, para avaliar a adequação da WDP-RT ao ambiente industrial, e em especial sua utilização por inspetores novatos.

Agradecimentos

Os autores agradecem a todos que participaram dos estudos experimentais e ao projeto SAUIM, ao CNPq, à FAPERJ, à FAPEAM e à Trópico Telecomunicações S.A. pelo apoio financeiro. Também agradecemos especialmente à empresa FabriQ Informática Ltda. pelo apoio no estudo de caso na indústria.

Referências Bibliográficas

- BASILI, V., ROMBACH, H., 1988. "The tame project: towards improvement-oriented software environments." IEEE Transactions on Software Engineering, v. 14, n. 6, pp. 758 - 773.
- BOLCHINI, D., GARZOTTO, F., 2007. "Quality of Web Usability Evaluation Methods: An Empirical Study on MiLE+". In: International Workshop on Web Usability and Accessibility (IWWUA) WISE 2007 Workshops, v. LNCS 4832, pp. 481 - 492, Nancy, France.
- CARVER, J., JACCHERI, L., MORASCA, S., SHULL, F., 2003. "Issues in Using Students in Empirical Studies in Software Engineering Education". In: Proceedings of the 9th International Symposium on Software Metrics (METRICS'03), pp. 239 – 249, Sydney.
- CONTE T., MASSOLAR, J., MENDES, E., TRAVASSOS, G., 2007. "Web Usability Inspection Technique Based on Design Perspectives". SBES 2007, João Pessoa, PB, Brasil.
- CONTE, T., MASSOLAR, J., MENDES, E., TRAVASSOS, P. G. H., 2009a. "Web Usability Inspection Technique Based on Design Perspectives." IET Software Journal, v. 3, n. 2, pp. 106-123.
- CONTE, T., VAZ, V., MASSOLAR, J., MENDES, E., TRAVASSOS, G. H., 2009b. "Improving a Web Usability Inspection Technique using Qualitative and Quantitative Data from an Observational Study". In: XXIII Simpósio Brasileiro de Engenharia de Software - SBES 2009, v. 1, pp. 227 - 235, Fortaleza, CE
- DENGER, C., KOLB, R., 2006. "Testing and inspecting reusable product line components: first empirical results". In: Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering (ISESE 2006), pp. 184 – 193, Rio de Janeiro.
- FERREIRA, S. B. L., LEITE, J. C. S. P., 2003. "Avaliação da Usabilidade em Sistemas de Informação: O Caso do Sistema Submarino". Revista de Administração Contemporânea - RAC, Curitiba, PR, v. 7, n. 3, p. 115-136.
- GOMES, M., CONTE, T., 2009. "WDP-RT: Uma técnica de leitura para inspeções de usabilidade de aplicações Web". Relatório técnico RT-DCC-ES001/2009. DCC/ UFAM.

- GOMES, M., CONTE, T., 2009b. "WDP-RT v2: Uma técnica de leitura para inspeções de usabilidade de aplicações Web". Relatório técnico RT-DCC-ES002/2009. DCC/ UFAM.
- GOMES, M., SANTOS, D. V., CHAVES, L., CASTRO, A., VAZ, V. T., SOARES, A., TRAVASSOS, G. H., CONTE, T., 2009. "WDP-RT: Uma técnica de leitura para inspeção de usabilidade de aplicações Web". In: VI Experimental Software Engineering Latin American Workshop (ESELAW 2009), v. 1, pp. 124-133, São Carlos, São Paulo.
- JURISTO, N., MORENO, A., SANCHEZ-SEGURA, M.-I., 2007. "Guidelines for Eliciting Usability Functionalities". IEEE Transactions on Software Engineering, v. 33, n. 11, pp. 744-758.
- KAPPEL, G., PRÖLL, B., REICH, S., RETSCHITZEGGER, W., 2006. "An Introduction to Web Engineering". In: Kappel, G., Pröll, B., Reich, S., Retschitzegger, W. Web Engineering: The Discipline of Systematic Development of Web Applications, John Wiley & Sons.
- MAFRA, S., BARCELOS, R., TRAVASSOS, G. H., 2006. "Aplicando uma metodologia Baseada em Evidência na Definição de Novas Tecnologias de Software". In: Proceedings of the 20th Brazilian Symposium on Software Engineering (SBES 2006), v. 1, pp. 239 – 254.
- MARCUS, A., 2002. "Return on Investment for Usable User-Interface Design: Examples and Statistics", Disponível online em: http://www.amanda.com/resources/ROI/AMA_ROIWhitePaper_28Feb02.pdf (último acesso: fevereiro de 2010).
- MEIRA, A., DA SILVA, F., 2009. "Abilities and Behavioural Profiles of SQA Professionals Related to Process Maturity Levels". In: XXIII Simpósio Brasileiro de Engenharia de Software - SBES 2009, 2009, Fortaleza. Anais do XXIII Simpósio Brasileiro de Engenharia de Software, 2009., v. 1, pp. 136 – 144, Fortaleza.
- MENDES, E., 2005. "A systematic review of Web engineering research." Proceedings of International Symposium on Empirical Software Engineering (ISESE 2005), IEEE Computer Science.
- MENDES, E., MOSLEY, N., COUNSELL, S., 2006. "The Need for Web Engineering: An Introduction". In: Mendes, E., Mosley, N. (eds), Web Engineering, Chapter 1, New York, Spinger Verlag.
- NIELSEN, J., 1993. "Usability Engineering". Academic Press, Cambridge, MA.
- NIELSEN, J., 1994. "Heuristic evaluation". In: Jacob Nielsen, Mack, R. L. (eds), Usability inspection methods, Heuristic Evaluation, New York, NY, John Wiley & Sons, Inc.
- NIELSEN, J., 1997. "Loyalty on the web". Alertbox, August, 1–3. Disponível online em <http://www.useit.com/alertbox/9708a.html> (último acesso: fevereiro de 2010).
- NIELSEN, J., 1999. "Design Web Usability". New Riders Publish, Indianapolis, Indiana, USA.
- PRATES, R. O., BARBOSA, S. D. J., (2003). "Avaliação de Interfaces de Usuário - Conceitos e Métodos". In: Coello, J. M. A., Fabbri, S. C. P. F. (eds), Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação, Capítulo 6, Campinas.
- SHULL, F., CARVER, J., TRAVASSOS, G. H., 2001. "An empirical methodology for introducing software processes". ACM SIGSOFT Software Engineering Notes, v. 26, n. 5, pp. 288-296.
- TRAVASSOS, G. H. (2008). Experimentação em Engenharia de Software: produzindo evidências, desenvolvendo tecnologias e melhorando processos de software (Apresentação Keynote). VII Simpósio Brasileiro de Qualidade de Software (SBQS 2008). Florianópolis.
- TRAVASSOS, G. H., BARROS, M., 2003. "Contributions of In Virtuo and In Silico Experiments for the Future of Empirical Studies in Software Engineering". In: Proceedings of the 2nd Workshop in Workshop Series on Empirical Software Engineering (WSESE 2003), pp. 117-130, Rome.
- TRAVASSOS, G. H., SHULL, F., FREDERICKS, M., BASILI, V., 1999. "Detecting defects in object-oriented designs: using reading techniques to increase software quality." ACM SIGPLAN Notices, v. 34, n. 10, pp. 47-56.