

Experimento – BISTWS e Metadados

1 - Definição do experimento

1.1 - Objeto de estudo: abordagem MTxTM (More Testable by Test Metadata), que é a abordagem BISTWS (Built-in Structural Testing of Web Services) com o uso de metadados (Eler et. al, 2010).

1.2 - Propósito: avaliar a utilização da abordagem MTxTM para testar serviços no contexto de uma composição. O objetivo é comparar os resultados do uso da abordagem MTxTM com o uso de uma abordagem baseada apenas em teste funcional (TF).

1.3 - Foco quantitativo: O foco primário é a cobertura estrutural obtida pela execução dos casos de teste criados em ambas as abordagens (TF e MTxTM) considerando os critérios todos-nós, todas-arestas e todos-usos. O foco secundário é a quantidade de casos de teste criados em cada abordagem e o número de falhas encontradas.

1.4 - Perspectiva: sob o ponto de vista de integradores de serviços. Os integradores têm acesso ao código-fonte da composição, mas não têm acesso ao código-fonte dos serviços utilizados por ela. O integrador deseja testar os serviços utilizados no contexto de uma composição, ou seja, os casos de teste serão criados para testar a composição, que invocará as operações dos serviços por ela utilizados.

1.5 - Contexto: o experimento será realizado por cerca de onze (10) alunos de pós-graduação do ICMC-USP. O experimento será realizado em um ambiente controlado (LabES - Laboratório de Engenharia de Software do ICMC-USP) e com tempo controlado.

2 - Planejamento do Experimento

2.1 - Definição das Hipóteses: a seguir são definidas as hipóteses nulas e as hipóteses alternativas deste experimento.

Hipóteses nulas:

- HN1 – Não existe diferença entre a cobertura dos requisitos de teste estrutural obtida pelos integradores que utilizam TF e a cobertura obtida pelos integradores que utilizam a abordagem MTxTM. A seguinte equação representa esta hipótese: $COV_{TF} = COV_{MTxTM}$
- HN2 – Não existe diferença entre a quantidade de falhas encontradas pelos integradores que utilizam TF e a quantidade de falhas encontradas pelos integradores que utilizam a abordagem MTxTM. A seguinte equação representa esta hipótese: $DEF_{TF} = DEF_{MTxTM}$

Hipóteses alternativas:

- HA1 – A cobertura dos requisitos de teste alcançadas pelos integradores que TF é menor do que cobertura obtida pelos integradores que utilizam a abordagem MTxTM. A seguinte equação representa esta hipótese: $COV_{TF} < COV_{MTxTM}$
- HA2 – A quantidade de falhas encontradas pelos integradores que utilizam TF é menor do que o número de falhas obtidas pelos integradores que utilizam a abordagem MTxTM. A seguinte equação representa esta hipótese: $DEF_{TF} < DEF_{MTxTM}$

2.2 – Objetos de estudo: neste experimento será utilizado um web service simples chamado de WS-GolfReservation (WS1) e uma composição chamada WS-GolfChampionship que utiliza o web

service WS-GolfReservation. . Os objetos utilizados são reais e foram obtidos em repositórios de software de código-aberto.

2.3 - Projeto do Experimento: os sujeitos do experimento utilizarão a abordagem TF (controle) ou a abordagem MTxTM (tratamento) para testar o web service WS-GolfReservation a partir do contexto da composição WS-GolfChampionship.

Os sujeitos que utilizarem a abordagem MTxTM terão acesso à análise de cobertura estrutural de WS-GolfReservation e dos metadados produzidos pelo desenvolvedor do serviço, enquanto que os sujeitos que utilizarem a abordagem TF terão acesso apenas à interface funcional de WS-GolfReservation.

Os sujeitos receberão treinamento adequado para cada uma das tarefas. Todos os sujeitos receberão treinamento para a criação de casos de teste para a composição WS-GolfReservation utilizando o framework JUnit.

Os sujeitos que utilizarem a abordagem MTxTM receberão instruções adicionais sobre como utilizar a abordagem para obter a análise de cobertura do serviço testável utilizado pela composição testada e também como utilizar os metadados para criar mais casos de teste e aumentar a cobertura obtida caso seja necessário.

Na Tabela 1 é apresentado o projeto de atribuição da abordagem que será utilizada pelos sujeitos do experimento. Os sujeitos do experimento serão divididos em dois grupos homogêneos em relação à experiência com a atividade de teste de software e a atribuição da abordagem que cada sujeito utilizará no quase-experimento será feita de forma aleatória.

Tabela 1: Projeto do experimento

Etapas\Sujeitos	TF	MTxTM
Sujeito 1	X	
Sujeito 2		X
Sujeito 3	X	
Sujeito 4		X
Sujeito 5	X	
Sujeito 6		X
Sujeito 7	X	
Sujeito 8		X
Sujeito 9	X	
Sujeito 10		X

2.4 - Seleção dos sujeitos: a seleção dos sujeitos do experimento será feita por conveniência entre alunos do grupo de pesquisa em engenharia de software.

2.5 - Seleção das variáveis: a seguir são apresentadas as variáveis independentes e dependentes deste experimento.

Variáveis independentes

- Abordagem MTxTM (tratamento)
- Abordagem TF (controle)
- WS-GolfReservation (objeto experimental)
- WS-GolfChampionship (objeto experimental)
- Experiência dos sujeitos na area de teste de software (contexto)

Variáveis dependentes

- A cobertura dos requisitos de teste obtida referentes aos critérios todos-nós, todas-arestas e todos-usos.

- O número de falhas encontradas. Os tipos de falhas serão coletados mas não farão parte da análise para o teste de hipótese.

2.6 – Descrição da instrumentação: o material fornecido aos sujeitos para a realização do experimento será composto de:

- Diretrizes para a realização do experimento.
- Código fonte e descrição textual das composições.
- Especificação textual do web service WS-GolfReservation (WSDL, objetivo geral, detalhamento de cada operação) e a especificação textual e o código-fonte da aplicação WS-GolfChampionship.
- Formulários para coleta de dados para cada tipo de abordagem utilizada.
- Formulário de pesquisa sobre o perfil (experiência com a área de teste de software) dos sujeitos.

2.7 - Ameaças à validade dos resultados do experimento: a seguir são apresentadas as ameaças à validade interna, de construção e externa deste experimento.

Validade interna: as ameaças à validade interna do experimento são as seguintes:

- *Experiência do aluno:* a experiência dos alunos com a utilização de web services e com a área de teste de software são fatores que podem influenciar os resultados do experimento. Neste experimento este fator será mitigado por meio de treinamento e por meio do balanceamento da experiência dos alunos na atribuição das abordagens utilizadas.
- *Serviços utilizados:* a complexidade dos serviços utilizados pode influenciar o resultado. Para mitigar este fator serão utilizados serviços de baixa complexidade desenvolvidos por terceiros e obtidos na internet. O serviço utilizado não foi desenvolvido pelos condutores do experimento para não haver o risco de que os serviços sejam criados para favorecer os resultados da abordagem avaliada.
- *Produtividade sob avaliação:* Pessoas tendem a ser mais produtivas quando estão sob avaliação. Esse fator pode influenciar os resultados porque em geral alunos que realizam experimentos tendem a pensar que estão sendo avaliados pelos resultados e não pela participação. Neste caso o experimento não será realizado no contexto de uma disciplina, mas serão voluntários que desejam colaborar com a pesquisa realizada em seu laboratório de pesquisa.

Validade de construção: As expectativas dos participantes ou dos condutores do experimento podem influenciar os resultados. Os participantes podem ter a tendência ou serem influenciados pelos condutores para satisfazer as hipóteses. Neste experimento os participantes não saberão das hipóteses para que este risco seja mitigado.

Validade externa: A população dos participantes não é representativa estatisticamente pois se trata de um grupo homogêneo de alunos de pós-graduação em Engenharia de Software. Além disso, a aplicação do experimento será feita em ambiente controlado e não em ambiente de produção real, embora os serviços utilizados sejam desenvolvidos por terceiros e utilizados em ambientes reais. Por esses motivos a validade externa deste quase-experimento fica comprometida e os resultados e conclusões não podem ser generalizados.

Validade de conclusão: a quantidade de amostras coletadas e o tipo de teste estatístico utilizado podem não ser adequados para o tipo de experimento realizado e invalidar as conclusões. Neste quase-experimento as conclusões serão obtidas pela utilização do teste T recomendado por Wohlin (referencia) para o tipo de projeto de experimento aqui utilizado. Desta forma, as conclusões obtidas pelo teste de hipótese deste quase-experimento serão válidas.

3 - Operação do experimento

3.1 – Preparação: os condutores do experimento farão contato com alunos de pós-graduação da área de Engenharia de Software para convidá-los a participar voluntariamente do experimento. Cada voluntário deverá preencher uma ficha para que os condutores do experimento possam identificar o perfil e a experiência de cada participante na área de teste de software.

Antes da execução do experimento os sujeitos receberão treinamento adequado para utilizarem a abordagem que será utilizada. O treinamento será realizado individualmente e imediatamente antes da execução do experimento e terá a duração de trinta minutos.

As diretrizes do experimento serão criadas e impressas para ser entregues a cada participante, assim como os formulários para a coleta de dados.

3.2 – Participantes: serão convidados alunos de pós-graduação de Engenharia de Software ICMC-USP e espera-se a participação de cerca de 10 alunos.

3.3 – Execução: este experimento será realizado entre os dias 12 e 19 de setembro de 2011 no Laboratório de Engenharia de Software ou nas dependências do ICMC-USP.

Os sujeitos executarão o experimento individualmente sob a supervisão dos condutores. Optou-se pela execução individual para aumentar o controle sobre o processo e para que o risco de falhas na execução seja menor. Se alguma falha for encontrada no início do experimento será possível corrigi-la para que os dados coletados dos outros participantes não sejam invalidados.

Os sujeitos terão no máximo 4 horas para realizar as atividades designadas. Os participantes deverão criar casos de teste para testar a composição WS-GolfChampionship com o objetivo de exercitar a integração e achar falhas em WS-GolfReservation, que é o serviço utilizado pela composição.

Para criar os casos de teste para a composição os sujeitos terão acesso a uma especificação e a uma descrição informal de WS-GolfChampionship, além do código-fonte. Para criar casos de teste com o objetivo de exercitar a integração entre WS-GolfChampionship e WS-GolfReservation, os participantes que utilizarem a abordagem TF terão acesso apenas à interface funcional de WS-GolfReservation e a uma descrição informal. Os sujeitos que utilizarem a abordagem MTxTM terão acesso a uma descrição informal, à interface funcional, a uma interface de teste que fornece a cobertura estrutural obtida pela execução de casos de teste e a uma lista de metadados de teste (a priori e on-demand) fornecida pelo serviço.

Os sujeitos deverão anotar na ficha de coleta de dados a quantidade de casos de teste criados e a quantidade de falhas encontradas. Além disso os sujeitos devem fornecer aos condutores do experimento os casos de teste criados. Os sujeitos que utilizarem a abordagem MTxTM ainda deverão anotar na ficha de coleta de dados a cobertura estrutural obtida de WS-GolfReservation com os casos de teste criados para WS-GolfChampionship.

3.4 - Validação dos dados: os formulários preenchidos pelos participantes serão conferidos para verificar se foram preenchidos corretamente.

4 - Análise e interpretação dos resultados

(omitido)

EXECUÇÃO DO EXPERIMENTO

Descrição da operação do experimento

O experimento foi realizado entre os dias 12 e 16 de setembro de 2011 com 10 alunos de pós-graduação do LabES/ICMC/USP. Uma parte dos sujeitos realizou o experimento na biblioteca do ICMC enquanto a outra parte realizou o experimento na sala de estudos do 2º andar do CISC e no LabES.

Os sujeitos executaram o experimento individualmente. Antes de cada execução o condutor do experimento realizou um treinamento de aproximadamente trinta minutos para cada sujeito. No treinamento o condutor explicou a cada sujeito o contexto no qual a aplicação a ser testada estava inserida e as características/especificações tanto da aplicação quanto do serviço por ela utilizado. O condutor também explicou de que forma os casos de teste deveriam ser criados e em qual formato (JUnit). Os sujeitos também receberam acesso ao código-fonte da aplicação a ser testada e também puderam acessar a base de dados do serviço utilizado pelo serviço.

Os sujeitos que utilizaram apenas a técnica de teste funcional foram orientados a criar casos de teste com base em critérios tais como classes de equivalência e análise de valor-limite. Os sujeitos que utilizaram a técnica MTxTM foram orientados a utilizar a técnica funcional no início e em seguida utilizarem as sugestões de casos de teste fornecidos pelo serviço para criar novos casos de teste e ampliar a cobertura estrutural obtida.

Durante a execução do experimento os sujeitos tiveram liberdade para perguntar ao condutor questões técnicas e de entendimento geral sobre a aplicação testada. Alguns sujeitos com menor experiência na criação de casos de teste com o JUnit precisaram de apoio para entender como criar as assertivas dos casos de teste.

Dois sujeitos que utilizaram a técnica MTxTM tiveram problemas na obtenção da análise de cobertura e das sugestões de casos de teste do serviço WS-GolfReservation porque o serviço estava com falhas e precisou ser substituído. A correção foi feita rapidamente e não prejudicou o andamento e nem o desempenho dos sujeitos.

Coleta de dados

Os sujeitos deveriam preencher todos os dados das fichas de coleta de dados para entregar ao condutor ao final do experimento, mas eles foram orientados a preencher apenas as informações básicas porque as outras informações podiam ser obtidas pela execução dos casos de teste e exigiam um esforço maior para serem fornecidas por meio dos formulários, como, por exemplo, a cobertura estrutural obtida para cada operação do serviço e para cada um dos critérios estruturais utilizados: todos-nós, todas-arestas e todos-usos.

Os sujeitos que utilizaram apenas a técnica funcional não tinham a opção de preencher as cobertura estrutural obtida porque eles não tinham acesso à interface de teste do serviço. Essas informações foram obtidas pelo condutor após o encerramento do experimento.

O condutor do experimento realizou uma análise detalhada dos casos de teste criados pelos sujeitos para identificar as falhas encontradas. Em alguns casos o número de casos de teste que falharam não correspondia ao número de falhas encontradas porque vários casos de teste se referiam à mesma falha. Neste caso foi contabilizada apenas uma falha. Houve situações em que os sujeitos criaram casos de teste que falharam, porém o erro estava na especificação do caso de teste. Um sujeito, por exemplo, criou um caso de teste para localizar uma partida criada em uma data e horário específico. O caso de teste falhou porque era esperado que a aplicação localizasse a partida mas a partida não foi encontrada. Ao analisar os casos de teste foi constatado que a partida procurada nunca havia sido criada e que o correto era que a aplicação não encontrasse a partida porque ela não existia. Esses casos não foram considerados como falha.

Dados coletados

Resultados da técnica MTxTM					
Sujeitos	Experiência	Cobertura	Total Falhas	#C.T.	Tempo
Sujeito 2	Alta	85	3	55	3h20min
Sujeito 4	Média	82	3	34	2h25min
Sujeito 6	Média	85	5	33	3h05min
Sujeito 8	Baixa	85	3	56	2h20min
Sujeito 10	Média	85	6	36	2h35min

Resultados da técnica funcional (TF)					
Sujeitos	Experiência	Cobertura	Total Falhas	#C.T.	Tempo
Sujeito 1	Média	74	5	36	2h20min
Sujeito 3	Alta	37	3	14	1h30min
Sujeito 5	Média	75	2	41	2h00min
Sujeito 7	Média	81	5	26	1h40min
Sujeito 9	Alta	72	2	14	1h55min

Trabalho da disciplina de Engenharia de Software Experimental

De acordo com o planejamento do experimento e dos dados coletados aqui apresentados pede-se o seguinte:

- Faça a análise estatística descritiva dos dados coletados e o teste estatístico das hipóteses do experimento, justificando a aceitação ou rejeição das duas hipóteses nulas. Considerar 5% de significância.
- Faça uma análise da correlação dos resultados do experimento com a experiência dos sujeitos e com o número de caso de teste criado por cada sujeito. Este ponto não é o objetivo fundamental do trabalho, mas será considerado como um esforço adicional.

O prazo inicial para a entrega deste trabalho é 03/10/2011, pelo site da disciplina do Moodle.