

# Experimento Piloto da abordagem de monitoração TWSM

## 1 - Definição do experimento

**1.1 - Objeto de estudo:** abordagem de monitoração TWSM (Testable Web Services Monitoring).

**1.2 - Objetivos:** avaliar a utilização da abordagem de monitoração TWSM para identificação de manutenções realizadas em web services estruturalmente testáveis desenvolvidos e fornecidos por terceiros. O estudo a ser realizado levará em conta apenas os requisitos funcionais dos web services e não considerará requisitos não-funcionais (SLAs).

**1.3 - Foco quantitativo primário:** O foco quantitativo primário é medir a precisão com a qual um monitorador consegue identificar as mudanças sofridas por um web service. O foco secundário é descobrir quais tipos de mudanças não são identificáveis pela abordagem TWSM e qual o número de casos de teste criado pelo monitorador após o início das atividades de monitoração.

**1.4 - Foco qualitativo:** O foco qualitativo primário é identificar a confiabilidade dos monitoradores nos resultados obtidos durante da atividade de monitoração com o método TWSM. O foco secundário é avaliar a qualidade dos casos de teste criados pelo monitorador durante a atividade de monitoração.

**1.5 - Perspectiva:** Monitoradores de web services adquiridos de terceiros. Os monitoradores são consumidores que desejam monitorar web services para perceber mudanças ocorridas através do tempo.

**1.6 - Contexto:** o experimento piloto será realizado por alunos de pós-graduação da área de Engenharia de Software do ICMC-USP.

## 2 - Planejamento do Experimento

### 2.1 - Definição das Hipóteses

#### Hipóteses nulas:

- HN1 – Seja  $M$  a quantidade de manutenções sofridas por um web service em um período de tempo  $t$ . Se  $M_{t_{\text{TWSM}}}$  for a quantidade de manutenções do web service identificadas pela abordagem de monitoração TWSM no período  $t$  e  $M_{\text{ad-hoc}}$  a quantidade de manutenções do web service identificadas no período  $t$  por uma abordagem de monitoração ad-hoc, então  $M_{t_{\text{TWSM}}} \leq M_{\text{ad-hoc}} \leq M$ . Isto significa que a identificação de manutenções realizadas pela abordagem ad-hoc é sempre mais precisa ou igual à identificação de TWSM.
- HN2 – Se  $OP_{t_{\text{TWSM}}}$  for um número que expressa a precisão com que o método TWSM identificou se cada operação do web service monitorado mudou ou não a cada atividade de monitoração, e  $OP_{\text{ad-hoc}}$  for um número que expressa essa precisão para a abordagem ad-hoc, então  $OP_{t_{\text{TWSM}}} \leq OP_{\text{ad-hoc}}$ .
- HN3 – Se  $GC_{t_{\text{TWSM}}}$  for o grau de confiança do monitorador que utiliza a abordagem TWSM de que as alterações do web service monitorado foram identificadas e se  $GC_{\text{ad-hoc}}$  for o grau de confiança do monitorador que utiliza uma abordagem ad-hoc, então  $GC_{\text{ad-hoc}} \geq GC_{t_{\text{TWSM}}}$ .

#### Hipóteses alternativas:

- HA1 – Seja  $M$  a quantidade de manutenções sofridas por um web service em um período de tempo  $t$ . Se  $M_{t_{\text{TWSM}}}$  for a quantidade de manutenções do web service identificadas pela abordagem de monitoração TWSM no período  $t$  e  $M_{\text{ad-hoc}}$  a quantidade de manutenções do web service identificadas no período  $t$  por uma abordagem de monitoração ad-hoc, então  $M_{\text{ad-hoc}} < M_{t_{\text{TWSM}}} \leq M$ . Isto significa que a identificação de manutenções realizadas pela abordagem TWSM é sempre mais precisa ou igual à identificação realizada pela abordagem ad-hoc.
- HA2 – Se  $OP_{t_{\text{TWSM}}}$  for um número que expressa a precisão com que o método TWSM identificou se cada operação do web service monitorado mudou ou não a cada atividade de monitoração, e  $OP_{t_{\text{TWSM}}}$  for um número que expressa essa precisão mas para a abordagem ad-hoc, então  $OP_{t_{\text{TWSM}}} \leq OP_{\text{ad-hoc}}$ .
- HA3 – Se  $GC_{t_{\text{TWSM}}}$  for o grau de confiança do monitorador que utiliza a abordagem TWSM de que as alterações do web service monitorado foram identificadas e se  $GC_{\text{ad-hoc}}$  for o grau de confiança do monitorador que utiliza uma abordagem ad-hoc, então  $GC_{t_{\text{TWSM}}} > GC_{\text{ad-hoc}}$ .

## 2.2 – Objetos de estudo

Os objetos de estudo deste experimento são dois web services desenvolvidos para a realização deste estudo. O primeiro é um web service chamado BankLoanMgrWS que oferece funcionalidades de empréstimos bancários. O segundo é um web service chamado IRWS que oferece funcionalidades de cálculos de Imposto de Renda segundo as regras definidas pela Receita Federal para ano de exercício 2010 e ano base 2009.

## 2.3 – Seleção de contexto

O processo deste experimento é on-line porque os participantes não realizarão o experimento em um momento específico mas durante um longo período. Os participantes são alunos de pós-graduação e não profissionais da área. Os objetos de estudo serão modelados especificamente para o experimento e por isso não são objetos reais. O experimento tem um contexto específico porque não pode ser generalizado para todas as situações.

## 2.4 – Projeto do Experimento

Neste experimento piloto um dos alunos fará o papel de fornecedor (F) e os outros farão o papel de monitorador (M1 e M2). O experimento durará duas semanas e a cada semana um web service diferente será monitorado (BankLoanMgrWS e IRWS). Os participantes não devem trocar informações sobre o experimento durante o seu período de execução.

Na primeira semana o fornecedor será responsável por realizar manutenções no web service BankLoanMgrWS. A cada manutenção o fornecedor deve publicar uma versão estruturalmente testável (interface funcional + interface de teste) e uma versão regular (apenas interface funcional) do web service. As alterações podem ser realizadas em qualquer horário do dia, mas as publicações das novas versões deverão ser obrigatoriamente realizadas sempre no período da tarde (12h – 18h). As alterações devem ser evolutivas, isto é, sempre serão feitas a partir da última versão modificada. O fornecedor não precisa fazer modificações todos os dias, mas precisa preencher os formulários de coleta de dados diariamente, mesmo que seja para indicar que nenhuma manutenção foi realizada.

Na primeira semana M1 usará uma abordagem ad-hoc e o monitorador M2 usará a técnica TWSM para monitorar BankLoanMgrWS. Os monitoradores deverão realizar as atividades de monitoração apenas uma vez entre os períodos de possível publicação de uma nova versão (entre 18h-0h ou 0h-12h).

Na segunda semana o mesmo procedimento da primeira semana será adotado. A diferença é que o web service mantido será o IRWS e M1 usará a abordagem TWSM enquanto M2 usará uma abordagem ad-hoc.

Os alunos participantes do experimento receberão treinamento adequado para cada uma das tarefas. O fornecedor receberá instruções sobre como alterar e publicar uma versão regular e uma versão testável do web service. Na primeira semana M1 receberá instruções para utilizar alguma técnica de monitoração ad-hoc e M2 receberá um treinamento para utilizar a abordagem de monitoração TWSM e a ferramenta de apoio. Na segunda semana os treinamentos serão invertidos.

## **2.5 – Seleção dos indivíduos (sujeitos)**

A seleção dos sujeitos do experimento piloto será feita por conveniência entre alunos do grupo de pesquisa em engenharia de software.

## **2.6 - Seleção das variáveis (fatores)**

### **Variáveis independentes:**

- Metodologia utilizada: uma abordagem ad-hoc baseada apenas em teste funcional (controle) e a abordagem TWSM para monitoração de web services estruturalmente testáveis (tratamento).
- Objetos de estudo: web services BankLoanMgrWS e IRWS.
- Casos de teste: os casos de teste fornecidos como conjunto de regressão para identificar mudanças ocorridas no web service monitorado.

### **Variáveis dependentes:**

- Número de manutenções identificado durante o período do experimento.
- Operações modificadas em cada manutenção.
- Grau de confiança de que as modificações foram devidamente identificadas por parte dos monitoradores.
- Casos de teste adicionados ao conjunto de regressão após o início do experimento.

### **Variáveis de contexto:**

- Conhecimento da linguagem Java e do framework JUnit.

## **2.7 – Descrição da instrumentação**

O material recebido pelo fornecedor (F) será composto de:

- diretrizes para a realização do experimento;
- código fonte Java de um web service e sua descrição;
- software de apoio para a geração de web services testáveis;
- formulário para coleta de dados sobre as manutenções realizadas.

O material recebido pelo monitorador usando o método adhoc será composto de:

- diretrizes para a realização do experimento;
- endereço de rede e descrição textual de um web service;
- um conjunto de casos de teste codificados em JUnit para testar o web service recebido.

O material recebido pelo monitorador usando o método TWSM será composto de:

- diretrizes para a realização do experimento;
- endereço de rede e descrição textual de um web service estruturalmente testável;

- uma ferramenta de apoio para testar web services estruturalmente testáveis;
- um conjunto de casos de teste codificados na ferramenta de apoio para testar o web service recebido;

Os materiais listados anteriormente estão nos anexos deste documento.

## 2.8 – Avaliação da validade

**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 o resultado da monitoração. No projeto piloto este fator será mitigado pela escolha de dois alunos com experiência na área de teste de software.
- Envolvimento do autor da abordagem. A participação do autor da abordagem no experimento representando qualquer um dos papéis poderá influenciar no resultado da monitoração. Se o autor for o fornecedor do web service poderá realizar alterações no web service que seriam mais facilmente percebidas por sua abordagem, mesmo que inconscientemente. A interferência no resultado seria menor se o autor participasse como monitorador, pois o fato de ter experiência na abordagem não muda o fato de não saber se e como as alterações foram feitas no web service monitorado. Para não influenciar diretamente os resultados do experimento o autor da abordagem fará apenas o papel de treinador.
- 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 e sim em um contexto colaborativo entre alunos de um mesmo laboratório de pesquisa.

**Validade de construção:** As ameaças à validade de construção do experimento são as seguintes:

- O experimento não reflete a imagem correta da teoria. Se o experimento não for uma representação do que ocorre na realidade os resultados podem não ser válidos. Neste experimento as alterações realizadas no web service são feitas artificialmente e não decorrente de uma mudança necessária por causa do ambiente, erros ou mudança de regras de negócio. Além disso, o web service monitorado não será observado no contexto de uma orquestração, mas isoladamente. Mesmo com esses fatores, acredita-se que as alterações sofridas pelo web service são manutenções que podem ocorrer na realidade e observar um web service isoladamente também ocorre em ambientes reais.
- 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:** é possível que os resultados do experimento não possam ser generalizados pelos seguintes fatores:

- A população dos participantes não ser representativa.
- A instrumentação não ser adequada à prática industrial.
- Período de aplicação do experimento. O experimento pode ser aplicado em um dia ou período especial que pode afetar os resultados.

## 3 - Operação do experimento

**Preparação:** os condutores do experimento farão contato com alunos de pós-graduação para convidá-los para participar voluntariamente do experimento. Um treinamento será criado para ser ministrado a cada participante. O treinamento para a abordagem ad-hoc terá duração de 30 minutos e o treinamento para a abordagem TWSM terá duração de 1 hora. As diretrizes do experimento serão criadas e impressas para serem entregues a cada participante, assim como formulários para identificar o perfil de cada um deles.

**Participantes:** serão convidados alunos de pós-graduação do ICMC-USP. O objetivo é contatar alunos com experiência na área de teste de software e com conhecimento da linguagem Java.

**Projeto do Experimento:** na Tabela 1 é apresentada a relação entre os sujeitos, abordagens e web services utilizados em cada etapa.

Tabela 1: Projeto do experimento piloto

Sujeitos	Semana 1: BankLoanMgrWS (WS1)	Semana 2: IRWS (WS2)
M1	Abordagem ad-hoc caixa-preta	Abordagem TWSM
M2	Abordagem TWSM	Abordagem ad-hoc caixa-preta

**Execução:** A realização do experimento será feita em duas etapas. Cada etapa durará exatamente uma semana. Antes de iniciar o experimento será atribuído a cada aluno participante um papel (fornecedor ou monitorador). Um aluno receberá o papel de fornecedor e os outros dois de monitoradores (M1 e M2). Antes de começa a primeira etapa, M1 receberá um treinamento sobre como monitorar um web service usando uma técnica ad-hoc baseada em teste funcional. M2 receberá um treinamento sobre como monitorar um web service usando a abordagem TWMS que utiliza teste funcional e estrutural. M2 também receberá um treinamento para poder utilizar a ferramenta de apoio à abordagem TWMS O fornecedor receberá um treinamento sobre os tipos de manutenções clássicas em Engenharia de Software e sobre como desenvolver e publicar um web service. O fornecedor também receberá treinamento sobre como gerar web services estruturalmente testáveis usando a abordagem BISTWS. Os participantes também serão instruídos sobre como preencher corretamente os formulários de coleta de dados do experimento de acordo com cada papel.

Na primeira semana o experimento será feito com base na monitoração do web service BankLoanMgrWS. O fornecedor deverá realizar manutenções no web service periodicamente e publicar uma versão testável e uma versão regular. As modificações e as publicações deverão sempre ser realizadas entre 12h e 18h. O fornecedor não precisa fazer alterações todos os dias, mas deve preencher os formulários de coleta de dados diariamente. No formulário o fornecedor deve informar a data e a hora de modificação do web service, o tempo gasto, se fez alguma modificação, quais as operações ou métodos internos alterados e qual o tipo de alteração realizada (refatoração, correção, mutação, etc).

Os monitoradores deverão realizar as atividades de monitoração apenas uma vez entre os períodos de possível manutenção do web service (entre 18h de um dia até 12h do outro dia). Os monitoradores deverão preencher o formulário a cada atividade de monitoração com as seguintes informações: data e hora, tempo gasto, mudanças identificadas, quais operações/métodos mudaram, a quantidade de casos de teste utilizada, os casos de teste utilizados, o recurso ou artefato que auxiliou a identificar a mudança (caso de teste, requisito de teste ou cobertura estrutural) e por fim o quanto ele está confiante de que a identificação realizada está correta.

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

**Análise e interpretação dos resultados primários**

Os formulários preenchidos pelos monitoradores serão comparados com os formulários preenchidos pelo fornecedor. Será verificado se os monitoradores conseguiram identificar corretamente quando ocorreu e quando não ocorreu mudança no web service. No caso de uma mudança ter sido identificada, será verificado, para cada abordagem, se o monitorador conseguiu identificar as partes modificadas. Também será analisada a confiança dos monitoradores em cada abordagem nas respostas dadas sobre a identificação da mudança.

Durante a análise dos dados serão feitas correlações entre as abordagens utilizadas, a quantidade de casos de teste gerados, os tipos de modificações identificadas e o tipo de recurso utilizado como indício de modificação (requisitos de teste, casos de teste ou cobertura estrutural).

Os resultados serão analisados pelo autor da abordagem com o auxílio de seu orientador acadêmico. Para cada hipótese, as análises dos dados serão feitas da forma mostrada a seguir.

### **H1 – Quantidade de manutenções identificadas**

Cada monitorador fará cinco (5) coletas durante uma semana. Cada coleta receberá uma pontuação no quesito identificação de manutenção. O formulário de coleta de dados do fornecedor será usado para atribuir pontuação a cada coleta da seguinte forma.

Para quando o fornecedor realizou modificação:

- O monitorador identificou modificação: um (1) ponto.
- O monitorador não identificou modificação: nenhum (0) ponto

Para quando o fornecedor não realizou modificação:

- O monitorador identificou modificação: nenhum (0) ponto.
- O monitorador não identificou modificação: um (1) ponto.

Para realizar a análise e verificar a validade da hipótese será assumido que  $M_{ad-hoc}$  é igual à soma dos pontos das coletas da abordagem ad-hoc e  $M_{TWSM}$  é igual à soma dos pontos das coletas da abordagem TWSM. O somatório dos pontos considerará as duas etapas do experimento.

### **H2 – Identificação das operações modificadas**

Cada monitorador fará cinco (5) coletas durante uma semana. Cada coleta receberá uma pontuação no quesito identificação da(s) operação(ões) alterada(s). O formulário de coleta de dados do fornecedor será usado para atribuir pontuação a cada coleta da seguinte forma:

- O monitorador identificou corretamente que uma operação foi alterada: dois (2) pontos.
- O monitorador identificou corretamente que nenhuma operação sofreu modificação: um (1) ponto.
- O monitorador identificou equivocadamente que uma operação foi alterada: nenhum (0) ponto
- O monitorador não identificou que uma operação foi alterada: menos um (-1) ponto.

Para realizar a análise e verificar a validade da hipótese será assumido que  $OP_{ad-hoc}$  é igual à soma dos pontos das coletas da abordagem ad-hoc e  $OP_{TWSM}$  é igual à soma dos pontos das coletas da abordagem TWSM. O somatório dos pontos considerará as duas etapas do experimento.

### **H3 – Grau de confiança na identificação da modificação**

Cada monitorador fará cinco (5) coletas durante uma semana. Cada coleta receberá uma pontuação no quesito confiança do monitorador de que as modificações foram devidamente identificadas.. O formulário de coleta de dados do fornecedor será usado para atribuir pontuação a cada coleta da seguinte forma, de acordo com o grau de confiança indicado.

- Grau de confiança alta: três (3) pontos.
- Grau de confiança média: dois (2) pontos.
- Grau de confiança baixa: um (1) ponto.

Para realizar a análise e verificar a validade da hipótese será assumido que  $GC_{ad-hoc}$  é igual à soma dos pontos das coletas da abordagem ad-hoc e  $GC_{TWSM}$  é igual à soma dos pontos das coletas da abordagem TWSM. O somatório dos pontos considerará as duas etapas do experimento.

## **Análise e interpretação dos resultados secundários**

### **Tipos de modificação**

O formulário preenchido pelo fornecedor será utilizado para identificar quais tipos de modificação foram detectados e quais não foram detectados pela técnica TWSM. Além disso, será feita a correlação entre tipo de modificação e tipo de recurso utilizado como indício (caso de teste, requisitos de teste ou análise de cobertura).

### **Casos de teste**

Os formulários preenchidos pelos monitoradores serão utilizados para identificar qual abordagem teve maior o número de casos de teste adicionais criados em cada iteração. Também será analisado se os casos de teste adicionais foram criados para identificar alguma alteração ou foram criados após a identificação de mudança.

Deseja-se avaliar se os casos de teste criados por quem usou a abordagem TWSM têm maior qualidade/eficácia do que os casos de teste criados por quem usou a abordagem ad-hoc. Deseja-se também identificar se os metadados e a interface de teste estrutural são úteis para auxiliar a derivação de novos casos de teste quando não se tem acesso ao código do web service.

