

Uma Investigação do Estado da Arte de Engenharia Reversa voltada para Testes de Software

Ricardo Castelo Branco¹, Taciana Novo Kudo¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74690-900 – Goiânia – GO – Brasil

{castelo.ricardo,taciana.inf}@gmail.com

Abstract. *Software maintenance is characterized as a big challenge for software engineering professionals. One of the difficulties is to test a software, to determine if it is according to your specifications, when the same documentation is out of date. In this sense, this paper proposes to identify how reverse engineering could be used to generate updated software documentation to support testing activities. It was performed a systematic review on works published between 2005 and 2014, aiming as main result the identification of research trends and gaps in that subject. Search results show a trend of reverse engineering being used to generate documentation for testing of graphical user interface (GUI).*

Resumo. *A manutenção de software caracteriza-se como um grande desafio para profissionais de Engenharia de Software. Uma das dificuldades é testar um software, para determinar se o mesmo está de acordo com suas especificações, quando a documentação do mesmo encontra-se desatualizada. Nesse sentido, este artigo propõe identificar como a engenharia reversa pode ser usada para gerar documentação atualizada do software de modo a apoiar as atividades de teste. Para tanto, foi realizada uma revisão sistemática sobre trabalhos publicados entre os anos de 2005 e 2014, almejando como resultado principal a identificação de tendências de pesquisa e lacunas no referido assunto. Resultados da pesquisa mostram uma tendência da engenharia reversa sendo usada para gerar documentação para testes de interface gráfica (GUI).*

1. Introdução

A fase de manutenção de software caracteriza-se como de grande desafio para profissionais de Engenharia de Software. Uma vez que o software encontra-se em uso, mantê-lo atualizado com as necessidades do cliente, sem perder a qualidade do mesmo é uma tarefa árdua. Soma-se a isso o fato de que manter a documentação desse software atualizada com todas as alterações e melhorias desenvolvidas diariamente não é a realidade da maioria das empresas.

Outra atividade que permeia a manutenção de software são os testes contínuos que precisam ser executados toda vez que o software sofre alguma alteração. Teste de software é o processo de execução de um software para determinar se ele está de acordo com suas especificações e se funcionou corretamente no ambiente para o qual foi projetado. Salienta-se que a atividade de teste é apontada como uma das mais onerosas no desenvolvimento e na manutenção de software.

A questão que motivou este trabalho é: Como testar um software para determinar se o mesmo está de acordo com suas especificações, se a sua documentação encontra-se desatualizada? Nesse contexto, a Engenharia Reversa [Penteado 1996] pode auxiliar na geração de documentação do software para recuperar informações do mesmo e produzir os modelos de análise e de projeto que servirão como base para execução dos testes de software.

O objetivo geral deste trabalho é identificar o estado da arte sobre Engenharia Reversa aplicada a teste de software. Como objetivos específicos, pretendem-se responder a determinadas questões de pesquisa, como por exemplo:

- Engenharia Reversa tem sido utilizada para apoiar a geração de documentação com a finalidade de executar testes de software?
- Nos últimos dez anos, tem-se observado crescimento da utilização de Engenharia Reversa para apoiar testes de software?
- Quais os principais artefatos de Engenharia Reversa que apoiam a execução de testes de software?

2. Metodologia

Nesta seção é discutida a metodologia utilizada para a realização deste trabalho, no caso, uma revisão sistemática, detalhando os passos realizados desde a preparação do protocolo, até a conclusão obtida a partir dos resultados da revisão sistemática.

2.1. Revisão Sistemática

Revisão sistemática é uma técnica de executar revisões abrangentes da literatura de forma não tendenciosa. [Pai et al. 2003]

Uma revisão sistemática sobre determinado tema possui o critério de seleção explícito, de forma que outros pesquisadores podem avaliar a qualidade da revisão e/ou executá-la novamente. Outra vantagem desta técnica é a produção de evidências sobre determinado tema, além da investigação de oportunidades de pesquisas nos desvios dos resultados encontrados.

A revisão sistemática [Sampaio and Mancini 2007] requer a definição de questões de pesquisa, a definição de uma estratégia de busca, o estabelecimento de critérios de inclusão e exclusão dos artigos e, uma análise criteriosa da qualidade da literatura selecionada. O processo de desenvolvimento desse tipo de estudo de revisão inclui caracterizar cada estudo selecionado, identificar conceitos importantes, comparar as análises estatísticas apresentadas e concluir sobre o que a literatura informa em relação a determinada intervenção, apontando ainda problemas/questões que necessitam de novos estudos.

Os principais passos da revisão sistemática proposta neste trabalho são:

1. Definição do protocolo da revisão sistemática: consiste em (a) definir as bases de conhecimento que serão utilizadas na revisão, como por exemplo, IEEEExplore e ACM DL; (b) definir as strings de busca que serão utilizadas para responder às questões de pesquisa apresentadas na seção de objetivos; (c) definir os critérios de inclusão e exclusão dos artigos encontrados;
2. Seleção de trabalhos: consiste em realizar as buscas de acordo com o protocolo que foi definido no passo anterior e, de acordo com os critérios de inclusão e exclusão, decidir quais artigos serão incluídos na atividade de extração de dados;

3. Extração e análise de dados: consiste na leitura e análise dos artigos de modo a identificar nos mesmos as respostas para as questões de pesquisa que foram definidas nos objetivos deste trabalho. Como resultado desta atividade será possível identificar o estado da arte de Engenharia Reversa para testes de software.

2.2. Processo da Revisão Sistemática

O processo seguido na realização desta revisão sistemática é composto de 5 fases, sendo elas: planejamento, execução, seleção, extração e conclusão.

2.2.1. Planejamento

A fase de planejamento é a fase inicial de uma revisão sistemática. Nesta fase nós definimos o protocolo que foi adotado em todas as fases subsequentes desta revisão, incluindo a formulação das questões de pesquisa, e a seleção das fontes e de estudos utilizados nesta pesquisa.

O foco do nosso estudo foi identificar o estado da arte da Engenharia Reversa de Software a fim de apoiar a atividade de teste de software. Com base neste objetivo, foram elaboradas seis questões relevantes para a pesquisa:

1. A engenharia reversa está sendo utilizada com o objetivo de apoiar teste de software?
2. Quais os principais artefatos de engenharia de software estão sendo produzidos para apoiar teste de software?
3. Os trabalhos que utilizaram engenharia reversa para fins de teste alcançaram resultados satisfatórios?
4. Existe algum método de engenharia reversa específico quando o objetivo é teste de software? Se não, quais os principais métodos utilizados?
5. Existe alguma ferramenta de engenharia reversa específica quando o objetivo é teste de software? Se não, quais as principais ferramentas utilizadas?
6. Para quais tipos de testes a engenharia reversa está sendo mais utilizada?

Para a realização de uma revisão sistemática de qualidade são necessárias fontes de pesquisas confiáveis. Partindo desse pressuposto, foram definidos alguns critérios para a seleção dessas fontes de pesquisa. Assim, elas deveriam:

- Fornecer mecanismos de busca por meio de strings de busca com suporte a expressões lógicas;
- Produzir os mesmos resultados sempre que a mesma string de busca for inserida;
- Disponibilizar mecanismo de consulta via web;
- Serem relacionadas a temas de Computação, incluindo Ciência da Computação, Sistemas de Informação, ou Engenharia de Software; e,
- Permitir filtro por ano de publicação.

Baseado nesses critérios, as fontes de pesquisa definidas foram:

- ACM Digital Library - <http://dl.acm.org/>
- Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) - <http://bdtd.ibict.br/>
- Engineering Village - <http://www.engineeringvillage.com/>

- IEEEEXplorer - <http://ieeexplore.ieee.org/search/advsearch.jsp?expression-builder>
- Science Direct - <http://www.sciencedirect.com/>
- Scopus - <http://www.scopus.com/>
- Springer - <http://link.springer.com/advanced-search>
- Web of Science - <http://apps.webofknowledge.com>

Ainda na fase de planejamento, foi definido que seriam considerados na revisão, os estudos escritos em Inglês ou Português, sendo estes trabalhos do tipo teses, dissertações, capítulos de livros com abstract e artigos de periódicos ou de conferências, incluindo estudos primários, secundários ou terciários, e com ano de publicação entre 2005 e 2014.

Após ter formulado as questões de pesquisa, estabelecido os critérios para a seleção das fontes em que seriam realizadas as pesquisas e, baseado nestes critérios, definidas as fontes de pesquisa, o próximo passo foi definir as strings de buscas que foram executadas em todas as fontes de pesquisa.

Foram definidas duas strings de busca, uma para pesquisar por trabalhos em inglês e outra para os trabalhos em português, ficando assim estabelecidas:

- String em inglês: ("reverse engineering"AND "testing software");
- String em português: ("engenharia reversa"AND "teste de software").

Finalmente, foram estabelecidos os critérios a serem usados para a exclusão dos estudos, sendo eles:

- Não escritos nos idiomas definidos neste protocolo;
- Trabalhos cujo texto completo não esteja disponível para acesso gratuitamente na Web ou no portal de periódicos da Capes; e,
- Trabalhos que não são relacionados com a área de testes de software nem com engenharia reversa.

2.2.2. Execução

Nessa fase foram executadas as strings de busca em cada uma das fontes de pesquisa definidas na fase de planejamento, além de preparar todas as informações retornadas.

Foi definido que os resultados retornados pelas fontes de pesquisa deveriam ser organizados em uma planilha eletrônica, a fim de facilitar a manipulação das informações. Para tanto, foi necessário exportar para um arquivo no formato Bibtex os resultados de cada busca efetuada nas fontes de pesquisa.

Após todos os resultados das fontes de pesquisa terem sido exportados para o formato Bibtex, centralizamos as informações dos trabalhos na planilha eletrônica.

2.2.3. Seleção

Nessa fase foram selecionados os trabalhos retornados nas buscas realizadas, conforme os critérios de inclusão e exclusão estabelecidos no protocolo.

A partir dessa seleção, foram lidos os títulos e os abstracts dos trabalhos, e atribuído o status de Incluído, Excluído ou Dúvida para cada artigo.

Para todo artigo marcado com o status de Excluído, deveria ser atribuído a ele um critério de exclusão, sendo eles:

- Não é um artigo, tese, dissertação ou capítulo de livro com abstract;
- Artigo duplicado em mais de uma base;
- Não trata especificamente de engenharia reversa, no entanto, dispõe sobre teste de software;
- Dispõe sobre engenharia reversa, mas não com foco em teste de software;
- Não dispõe sobre engenharia reversa, tampouco sobre teste de software.

Os artigos marcados com o status de Dúvida foram analisados pela professora orientadora e retificados com o status correto (incluído ou excluído).

2.2.4. Extração

Nessa fase foi realizada a leitura na íntegra de todos os artigos marcados com o status de Incluído na fase de seleção, a fim de extrair as seguintes informações:

- Qual artefato de engenharia de software está sendo gerado na engenharia reversa?
- Que tipo de notação é utilizado para fazer os artefatos de engenharia de software?
- Utiliza, cita ou propõe algum método de engenharia reversa?
- Utiliza, cita ou propõe alguma ferramenta específica de engenharia reversa?
- Qual tipo de teste de software é o foco do trabalho?
- Qual técnica de teste de software é citada ou utilizada?

2.2.5. Conclusão

Nessa última fase da revisão sistemática, foi possível emitir todos os relatórios e gráficos embasados nos resultados encontrados nas fontes de pesquisa.

3. Análise dos Resultados

Na fase de execução, foram retornados 510 artigos nas oito fontes de pesquisas, sendo que, a única base a não retornar artigos foi a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD).

Em seguida, na fase de seleção, aos trabalhos que retornaram das pesquisas realizadas, foi atribuído o status de Incluído, Excluído ou Dúvida, sendo esses últimos analisados pela professora orientadora e retificados com o status correto. Foram assinalados 22 artigos com o status de incluído, 452 excluídos e 36 como dúvida. Após a análise dos artigos marcados como dúvida, 8 foram incluídos e 28 excluídos.

Portanto, ao final da fase de seleção, 30 artigos foram incluídos e 480 excluídos.

Dentre as exclusões, podemos destacar que 22% foram de artigos duplicados entre as bases, o que é justificado pela utilização de diversas fontes de pesquisa, e 71% foram excluídos pelos critérios relacionados à string de busca, especialmente os artigos que não dispõem sobre engenharia reversa, tampouco sobre teste de software e os que tratam somente de teste de software.

Na fase de extração, a qual corresponde à etapa onde os dados relevantes à pesquisa podem ser extraídos, pode-se observar que 05 artigos tratavam somente de teste de software, 02 tratavam somente de engenharia reversa, e 01 artigo não tratava nem sobre engenharia reversa, e nem sobre teste de software.

Ademais, foi acrescentado um sexto critério de exclusão, relacionado aos artigos que não estavam disponíveis para download, totalizando 04 artigos nesta situação.

Finalmente, dos 30 trabalhos que iniciaram a fase de extração, 12 artigos foram excluídos, totalizando 18 artigos dos quais foi possível extrair algum tipo de informação relevante para a pesquisa.

É importante ressaltar que entre os anos de 2010 e 2014 foram selecionados 11 artigos, 57% a mais que o período compreendido entre os anos de 2005 e 2009, observando um expressivo aumento de interesse na área de pesquisa objeto deste trabalho.

Dos 17 artigos que apresentam artefatos de engenharia de software gerados a partir da engenharia reversa, pode-se destacar que 53% geraram diagramas de fluxo e 35% geraram um modelo estrutural de cada estado de interface gráfica.

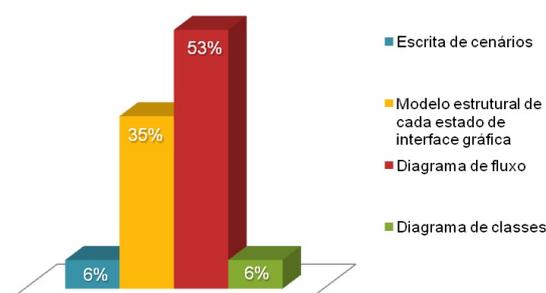


Figura 1. Artefatos de engenharia de software gerados na engenharia reversa.

Com relação à notação utilizada para elaborar estes artefatos, foi observado que em sua maioria, 59%, foi utilizado máquinas de estado e 23% dos 17 artigos utilizaram arquivos XML.

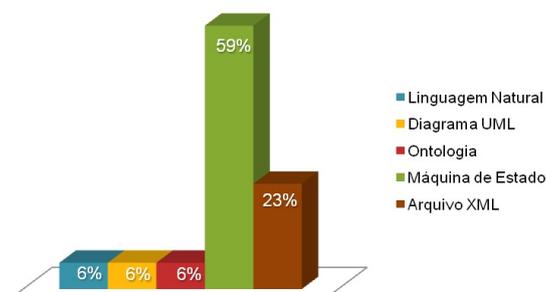


Figura 2. Notações utilizadas para elaborar os artefatos de engenharia de software.

Nos 18 artigos observamos as técnicas de teste utilizadas, 50% testaram a interface gráfica do software, enquanto 32% realizaram teste estrutural.

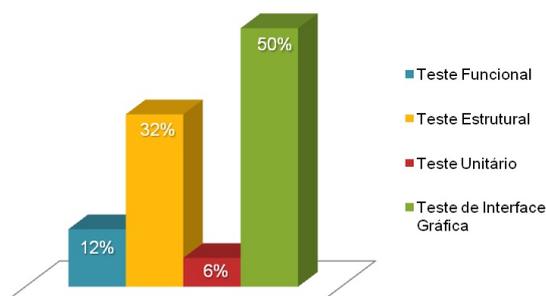


Figura 3. Técnicas de teste de software observados nos artigos.

Para finalizar a análise da fase de extração, podemos observar que:

- Com relação aos artefatos gerados na engenharia reversa, a maioria produz diagrama de fluxos ou modelo estrutural de cada estado de interface gráfica;
- As notações mais utilizadas são diagramas de máquina de estado e arquivos XML;
- Os tipos de testes apoiados por essas engenharias reversas foram teste de interface ou teste estrutural.

Conclui-se que, para realizar testes estruturais, quando não se tem documentação do software, é aplicada a engenharia reversa a fim de gerar os diagramas de fluxos, normalmente utilizando a notação de máquinas de estado. A partir do entendimento do software neste nível de estados, é possível realizar o teste estrutural.

Para os testes de interface, a engenharia reversa é utilizada para gerar modelo estrutural de cada estado da interface gráfica, normalmente com diagramas de máquina de estados ou arquivos XML. A partir desses artefatos, a interface pode começar a ser testada.

4. Considerações Finais

A proposta deste trabalho foi identificar o estado da arte sobre Engenharia Reversa aplicada a teste de software, pretendendo responder a determinadas questões de pesquisa, como por exemplo:

- Engenharia Reversa tem sido utilizada para apoiar a geração de documentação com a finalidade de executar testes de software?
- Nos últimos dez anos, tem-se observado crescimento da utilização de Engenharia Reversa para apoiar testes de software?
- Quais os principais artefatos de Engenharia Reversa que apoiam a execução de testes de software?

Para responder a essas questões de pesquisa, foi realizada uma revisão sistemática da literatura, que é um estudo secundário que tem por objetivo reunir, avaliar criticamente e apresentar um resumo dos resultados de múltiplos estudos, utilizando um protocolo que norteou a pesquisa e os métodos aplicados durante o trabalho.

Dos 510 artigos retornados inicialmente, um total de 18 foram possíveis extrair informações consideradas relevantes à pesquisa, obtendo as seguintes conclusões:

- A grande maioria dos artigos foi excluída na fase de seleção pelos critérios relacionados à string de busca, o que pode sugerir um problema em sua definição;
- As fontes de pesquisa IEEE e Scopus incluem a maioria dos artigos relevantes sobre o tema deste trabalho;
- Nos últimos cinco anos da pesquisa houve um expressivo aumento de publicações com o tema deste trabalho, o que demonstra um crescimento da utilização de Engenharia Reversa para apoiar testes de software;
- Todos os artigos que tiveram como foco o teste estrutural geraram diagramas de fluxo na engenharia reversa;
- Todos os artigos que geraram na engenharia reversa modelo estrutural de cada interface gráfica, fizeram a fim de testá-la;
- Apesar de não ficar evidenciado que existe algum método de engenharia reversa específico quando o objetivo é teste de software, podemos destacar o método de engenharia reversa dinâmica como o mais utilizado nos artigos.
- Finalmente, observou-se que a engenharia reversa é amplamente utilizada para realizar teste de interface gráfica.

Com os resultados obtidos na realização deste trabalho e a fim de dar continuidade a pesquisas sobre este assunto, são sugeridas uma pesquisa qualitativa relacionada ao uso da engenharia reversa na geração de modelo estrutural de interface gráfica, e outra relacionada ao teste de interface gráfica baseado em modelos.

Sugere-se ainda que, à medida que os programadores elevam a utilização da engenharia reversa com o objetivo de testar software, trabalhos futuros sejam realizados para analisar os mesmos pontos abordados nesta revisão sistemática.

Referências

- Pai, M., McCulloch, M., Gorman, J. D., Pai, N., Enanoria, W., Kennedy, G., Tharyan, P., and Colford Jr, J. M. (2003). Systematic reviews and meta-analyses: an illustrated, step-by-step guide. *The National medical journal of India*, 17(2):86–95.
- Penteadó, R. A. D. (1996). *Um método para engenharia reversa orientada a objetos*. Tese de Doutorado, Universidade de São Paulo, São Carlos-SP.
- Sampaio, R. F. and Mancini, M. C. (2007). Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Braz. J. Phys. Ther.(Impr.)*, 11(1):83–89.