Caracterização da qualidade interna de ferrramentas de análise estática de código fonte

Joenio Marques da Costa
Universidade Federal da Bahia (UFBA)
joenio@colivre.coop.br

23 de abril de 2016

1 Introdução

(à fazer)

1.1 Contribuições esperadas

(à fazer)

2 Fundamentação teórica

(à fazer)

3 Metodologia

Neste capítulo será apresentada a metodologia utilizada no estudo como meio de validar as seguintes hipóteses:

H1: *Existem publicações sobre ferramentas de análise estática com disponibilidade de código-fonte*

H2: *Existem ferramentas de análise estática disponíveis livremente na indústria com disponibilidade de código-fonte*

H3: *Existem valores de referência para métricas de código-fonte para ferramentas de análise estática*

H3: *Ferramentas da indústria possuem melhores valores de métricas de código-fonte*

As seções à seguir descrevem as atividades de cada etapa da metodologia.

3.1 Planejamento do estudo

3.1.1 Seleção das métricas

(à fazer)

3.1.2 Seleção das fontes de ferramentas de análise estática

Para ser possível validar as hipóteses aqui levantadas é necessário realizar uma busca por ferramentas de análise estática desenvolvidas no contexto da academia e da indústria, para isso, será feito um planejamento detalhado para realizar a seleção de ferramentas em cada um destes contextos.

Academia No contexto acadêmica a busca por ferramentas será feita através de artigos publicados em conferências que tenham histórico de publicação sobre ferramentas de análise estática de código fonte. Estes artigos serão analisados e aqueles com publicação de ferramenta de análise estática serão selecionados.

Indústria Na indústria a busca por ferramentas será feita a partir de referências encontradas na internet, algumas organizações mantém listas de ferramentas para análise de código-fonte, a Wikipedia também mantém uma lista de ferramentas, estas referências serão utilizadas como ponto de partida e cada ferramenta será analisada a fim de validar se são da indústria ou surgiram em contexto acadêmico.

Uma vez que as ferramentas tenham sido selecionadas inicia-se a extração de seus atributos de qualidade interna.

3.1.3 Seleção da ferramenta de análise estática de código-fonte

Para realizar a caracterização das ferramentas através dos seus atributos de qualidade interna é necessário uma ferramenta capaz de analisar estaticamente o código-fonte destas ferramentas e extrair atributos relacionados à sua qualidade interna. Para isto utilizaremos o Analizo(KON, 2010). Falta Justificar! Quais vantagens? Referencias?

3.2 Coleta de dados

A partir das fontes selecionadas na etapa anterior serão realizadas duas atividades para identificar e mapear as ferramentas de análise estática com código-fonte disponível, uma atividade relacionada ao levantamento de ferramentas da academia, outra atividade relacionada ao levantamento de ferramentas da indústria.

3.2.1 Ferramentas da academia

A seleção de ferramentas será relizada através de uma revisão estruturada dos artigos selecionados a partir das seguintes conferências:

- ASE - Automated Software Engineering[1]
- CSMR[2] - Conference on Software Maintenance and Reengineering[3]
- SCAM - Source Code Analysis and Manipulation Working Conference[4]

Chamamos de revisão estruturada um processo disciplinado para seleção de artigos a partir de critérios bem definidos de forma que seja possível a reprodução do estudo por parte de pesquisadores interessados. Alguns resultados preliminares podem ser consultados na Tabela 1 da Seção 4.1.

[1] http://ase-conferences.org
[2] A conferência CSMR tornou-se SANER - Software Analysis, Evolution, and Reengineering a partir da edição 2015.
[3] http://ansymore.uantwerpen.be/csmr-wcre
[4] http://www.ieee-scam.org

3.2.2 Ferramentas da indústria

Na indústria tomaremos como ponto de partida a lista de ferramentas de análise estática mantida pelo projeto SAMATE[5] - *Software Assurance Metrics and Tool Evaluation* disponível em NIST (2016), mais sobre o projeto SAMATE pode ser encontrado em Ribeiro (2015).

O site do software Spin mantém uma lista de ferramentas comerciais e de pesquisa para análise estática de código-fonte para C em Spin (2016).

O Instituto de Engenharia de Software do CERT mantém uma lista de ferramentas de análise estática em CERT (2016).

O site da ferramenta Flawfinder leva a um link com referências para inúmeras ferramentas livres, proprietárias, gratuitas mas não-livres de ferramentas de análise estática e outros tipos de análise em Wheeler (2015).

Uma outra fonte contendo uma relação extensa de ferramentas é mantida na Wikipedia em Wikipedia (2016).

3.3 Caracterização dos artigos

Caracterização dos papers analisados na revisão estruturada e caracterização teórica do ecosistema das ferramentas da academia.

3.4 Caracterização das ferramentas

Será realizada uma caracterização prática das ferramentas, tanto acadêmica quando da indústria, através da análise e extração de métricas de código-fonte das mesmas.

3.5 Exemplo de uso

Por fim, os valores de métricas de referência encontradas serão utilizadas como guia para refatorar a ferramenta Analizo.

(à fazer)

4 Conclusão

4.1 Resultados preliminares

A Tabela 1 apresenta um resumo do número de artigos em cada edição do SCAM e quantos artigos traz publicação de ferramenta de análise estática com código fonte disponível.

As Tabelas 3 e 2 apresentam as ferramentas do NIST após avaliação inicial sobre disponibilidade do código-fonte. Dos 54 apenas 19 tinham código fonte disponível e encontrável facilmente. Todos foram baixados em "dataset/NIST".

Assim, temos um total de 19 ferramentas da indústria com código-fonte disponível e ??? da academia com código fonte disponível, é preciso avaliar em qual linguagem de programação foi escrita cada ferramentas pois só iremos analisar aquelas em C, C++ o Java que são suportadas pelo Analizo.

[5] http://samate.nist.gov

Tabela 1: Total de artigos analisados por edições do SCAM

Edição	Total de artigos	Artigos com ferramenta
SCAM 2001	23	-
SCAM 2002	18	-
SCAM 2003	21	-
SCAM 2004	17	-
SCAM 2005	19	-
SCAM 2006	22	2
SCAM 2007	23	1
SCAM 2008	29	-
SCAM 2009	20	-
SCAM 2010	21	1
SCAM 2011	21	1
SCAM 2012	22	4
SCAM 2013	24	-
SCAM 2014	35	1
SCAM 2015	?? (pendente)	?
Total	315	10

Ferramenta utilizada para isto foi a sloccount, uma ferramenta livre para contagem de linhas de código fonte, dá estatística de em qual linguagem é escrita em porcentagem. Abaixo destaco a linguagem de programação que tem maior porção em porcentagem.

Após analise ficamos com um total de 25 ferramentas, 15 da indústria e 10 da academia.

Segue a lista de todos os projetos e qual a fonte (industria ou academia):

4.2 Cronograma

(à fazer)

Referências

CERT. *Secure Coding Tools*. 2016. [Online; acessado 23 Abril de 2016]. Disponível em: ⟨http://www.cert.org/secure-coding/tools/index.cfm⟩.

KON, A. T. J. C. J. M. P. M. L. R. R. L. A. C. C. F. Analizo: an extensible multi-language source code analysis and visualization toolkit. p. 6, 2010.

NIST. *SAMATE - Source Code Security Analyzers*. 2016. [Online; acessado 20 Abril de 2016]. Disponível em: ⟨http://samate.nist.gov/index.php/Source_Code_Security_Analyzers.html⟩.

RIBEIRO, A. C. Análise estática de código-fonte com foco em segurança: Metodologia para avaliação de ferramentas. 2015.

SPIN. *Static Source Code Analysis Tools for C*. 2016. [Online; acessado 23 Abril de 2016]. Disponível em: ⟨http://www.spinroot.com/static⟩.

Tabela 2: Lista de ferramentas do SAMATE - NIST com código fonte não disponível

Ferramenta	Avaliacao
ABASH	código não disponível
ApexSec Security Console	código não disponível
Astrée	código não disponível
bugScout	código não disponível
C/C++test®	código não disponível
dotTEST™	código não disponível
Jtest®	código não disponível
HP Code Advisor (cadvise)	código não disponível
Checkmarx CxSAST	código não disponível
CodeCenter	código não disponível
CodePeer	código não disponível
CodeSecure	site offline
CodeSonar	código não disponível
Coverity SAVE™	código não disponível
Csur	código não disponível
DoubleCheck	código não disponível
Fluid	código não disponível
Goanna Studio and Goanna Central	código não disponível
HP QAInspect	código não disponível
Insight	código não disponível
ObjectCenter	código não disponível
Parfait	código não disponível
PLSQLScanner 2008	código não disponível
PHP-Sat	link para código offline
PolySpace	código não disponível
PREfix and PREfast	código não disponivel
QA-C, QA-C++, QA-J	código não disponível
Qualitychecker	código não disponível
Rational AppScan Source Edition	código não disponível
Resource Standard Metrics (RSM)	código não disponível
SCA	código não disponível
SPARK tool set	código não disponível
TBmisra®, TBsecure®	código não disponível
PVS-Studio	código não disponível
xg++	código não disponível

Tabela 3: Lista de ferramentas do SAMATE - NIST com código fonte disponível

Ferramenta	Avaliacao
BOON	código disponível
Clang Static Analyzer	código disponível
Closure Compiler	código disponível
Cppcheck	código disponível
CQual	código disponível
FindBugs	código disponível
FindSecurityBugs	código disponível
Flawfinder	código disponível
Jlint	código disponível
LAPSE	código disponível
Pixy	código disponível
PMD	código disponível
pylint	codigo disponivel
RATS (Rough Auditing Tool for Security)	código disponível
Smatch	código disponível
Splint	código disponível
UNO	código disponível
Yasca	código disponível
WAP	código disponível

WHEELER, D. A. *Static analysis tools for security*. 2015. [Online; acessado 23 de Abril de 2016]. Disponível em: ⟨http://www.dwheeler.com/essays/static-analysis-tools.html⟩.

WIKIPEDIA. *List of tools for static code analysis*. 2016. [Online; acessado 23 Abril de 2016]. Disponível em: ⟨https://en.wikipedia.org/wiki/List_of_tools_for_static_code_analysis⟩.

Tabela 4: Lista com total de ferramentas a ser analisadas

Ferramenta	Linguagem	Fonte
BOON	ansic	industria
CQual	ansic	industria
RATS	ansic	industria
Smatch	ansic	industria
Splint	ansic	industria
UNO	ansic	industria
Clang Static Analyzer	cpp	industria
Cppcheck	cpp	industria
Jlint	cpp	industria
WAP	java	industria
Closure Compiler	java	industria
FindBugs	java	industria
FindSecurityBugs	java	industria
Pixy	java	industria
PMD	java	industria
Indus	java	academia
TACLE	java	academia
JastAdd	java	academia
WALA	java	academia
error-prone	java	academia
AccessAnalysis	java	academia
Bakar Alir	java/ada/python	academia
InputTracer	ansic	academia
srcML	cpp/cs (cs = C# ?)	academia
Source Meter	java	academia