Test Coverage Measurement and Analysis on the Basis of Software Traceability Approaches

Muhammad Shahid, Suhaimi Ibrahim, and Harihodin Selamat

Abstract—Test Coverage is an important step in software testing and a good indicator of software quality. It helps in evaluating the effectiveness of testing by providing data on different coverage items. It ensures that testing is effectively carried out without missing out all the key functional areas or the features of the software under development. We have analyzed several software traceability approaches to understand the existing test coverage framework. These approaches have been classified in different categories. Moreover, we have compared researchers work based on these approaches. Comparison has been made based on some criteria such as coverage types, data base support, software traceability support, regression testing support, coverage construction and coverage re-construction techniques. We are of the opinion, although much research effort has been put on how to get coverage information by either code based testing or requirement based testing, not much has been paid to measure the coverage by combining both to get integrated coverage.

Index Terms—Test Coverage, Code Coverage, Software Traceability, Software Testing.

I. INTRODUCTION

Software testing is an essential activity in software maintenance life cycle. It is a practice often used to determine and improve software quality. Testing activities include obtaining the test coverage. Coverage is the extent that a structure has been exercised as a percentage of the items being covered. If coverage is not 100%, then more tests may be designed to test those items that were missed and therefore, increase coverage [1]. Test coverage can help in monitoring the quality of testing, and assist in directing the test generators to create tests that cover areas that have not been tested before [2].

There are several test coverage analysis approaches that have been proposed from a very simple way until the latest ones that apply very complex techniques. This paper aims at evaluating several recent test coverage analysis approaches. The initial results obtained from this evaluation can be used to observe to what extent each approach has a capability to support test coverage analysis. Although much research attention has been given to Test Coverage measurement and analysis, but there is a need to cover all three granularity levels of test coverage items, i.e. fine-grain, medium-grain and course-grain in detail.

The remaining part of this paper is organized as follows. Section II describes the test coverage in software testing. It provides definition of test coverage and the usage of its analysis in different fields of software testing. It is followed by section III which offers the survey of recent test coverage approaches. Section IV provides evaluation of these approaches based on a set of criteria and finally section V concludes this paper.

II. TEST COVERAGE

Coverage is a quality insurance metric which determines how thoroughly a test suite exercises a given program. The output of coverage measurement can be used in several ways to improve the verification process. First, the coverage tool can detect illegal events that occur and help find bugs in the design. Coverage measurement can also provide the user with information on the status of the verification process. It can help to find holes in the testing, i.e. areas that are not covered [2]. Test coverage also helps in Regression testing, test case prioritization, test suite augmentation and test suite minimization.

A. Test Coverage Items

Based on the research by different researchers [15], there are three granularity levels of test coverage items, i.e. fine-grain, medium-grain and course-grain.

Fine grain level consists of Statement, Basic Block, Branch, Condition, Condition / Decision coverage, Modified Condition MCDC, and LCSAJ (Linear Code Sequence And Jump) items. Method or Function, Class, Package and Design items are categorised in medium-grain while Requirements are considered as course-grain item.

B. Test Coverage Process

The Test/code coverage analysis process is generally divided into three tasks: code instrumentation, data gathering, and coverage analysis. Code instrumentation consists of inserting some additional code to measure coverage results. Instrumentation can be done at the source level in a separate pre-processing phase or at runtime by instrumenting byte code. Data gathering consists of storing coverage data collected during test runtime. Coverage data analysis consists of analysing the collected results and providing test strategy recommendations in order to reduce, to feed or to modify the relevant test suite.

III. CLASSIFICATION OF TEST COVERAGE APPROACHES

This section will present recent state-of-art approaches that have produced solution for test coverage measurement
and analysis. We have identified some significant approaches that can contribute to the further development and enhancement of test coverage measurement and analysis process.

SOFTWARE TRACEABILITY APPROACHES

The researchers [3] divide traceability approaches into seven categories:

- Information retrieval
- Rule based
- Event based
- Hypertext-based
- Feature model-based
- Value-based
- Scenario-based

A. Information Retrieval Approach (IRA)

This approach focuses on automating the generation of traceability link by similarity comparison between two types of artifacts [4]. The two basic IR models which commonly used in traceability generation are probabilistic and vector space models. Numerous variant models have also been applied including the popular model Latent Semantic Indexing which is based on vector space model. In each model, one type of particular artifacts treats as a query and another type of artifacts treats as a document being searched in term of the query. For example, source code treats as a query against requirements specification as a document being searched based on the query.

The general steps include (i) preprocessing, i.e. stop word removal and or stemming, (ii) analyzing and indexing of an incoming document collection, followed by constructing a representation of each document and then archiving them, (iii) analyzing and representing incoming queries, and using a matching or ranking algorithm to determine which document representations are similar to the query representation. The scope of tracing covers almost of artifacts including high-level and low level requirements, manual documents, design elements, test cases, and source code.

B. Rule-Based Approach (RB)

RB covers requirement statement documents, use case documents, and analysis object models as the objects of tracing. Spanoudakis et al. in [5] propose a method to automatically create traceability link using rules. They use two traceability rules, i.e. requirement-to-object-model traceability rule (RTOM rule) and inter-requirement traceability rule (IREQ rule). The rules are deployed into three specific documents types, i.e. (i) requirements statement documents (RSD), (ii) use case documents (UCD), and (iii) analysis object models (AOM). RTOM rules are used to trace the RSD and UCD to an AOM, while IREQ rules are used to trace between RSD and the UCD. The method assumes that all of document types are in XML-based format. The traceability rules are also represented in XML-based markup language. The method consists of four stages, i.e. (i) grammatical tagging of the artifacts, (ii) converting the tagged artifacts into XML representations (iii) generating traceability relations between artifacts, and (iv) generating traceability relations between different parts of the artifacts.

C. Event-Based Approach (EB)

In this approach, Traceability relationships are defined as publisher-subscriber relationships in which dependent artifacts must subscribe to the requirements on which they depend. When a requirement change, the dependent artifacts are notified and subsequently proper action can be taken [6]. The method involves three main components, i.e. (i) the requirement manager which is responsible for managing requirements and for publishing change event messages to the event server, (ii) the event server which responsible for establishing traceability by handling initial subscriptions placed by dependent entities, and also listening for event notifications from the requirement manager(s) and forwarding event messages to relevant subscribers, and (iii) the subscriber manager which responsible for listening on behalf of the subscribers that it manages for event notifications forwarded by the event server.

D. Hypertext-Based Approach (HB)

Maletic et al. in [7, 8] proposed an approach. This approach utilizes XML as the main tool for representing models and created links. The models and their links are converted into XML-based representation. Models are categorized into anchor model and target model. The links are established between anchor and target model with particular link types, i.e. causal, non-causal, and or navigational links. Once the model-to-model traceability links have been established, meta-differencing mechanism is used to indicate if some changes have been occurred in the models. The evolution is supported by a fine-grained versioning technique. The scope of tracing includes all types of artifacts.

E. Feature Model-Based Approach (FB)

Riebisch and Pashov in [9, 10] describe feature model-based method for requirement traceability. They utilize feature modeling which describes a requirements as an overview and models the variability of a product line. A feature model consists of a graph with features as nodes and feature relations as edges. If the number of features is very high, then the representation of features and their relations are displayed by tables. FB is applied for the definition of a product by a customer. Every feature describes a property of a product from the customer’s point of view. There are three categories of features, i.e. (i) functional features (ii) interface features, and (iii) parameter features.

F. Value-Based Approach (VB)

VB consists of five processes, i.e. (i) requirements definition, that is identifying atomic requirements and assigning an identifier to each of them, (ii) requirements prioritization, that is estimating the value, risk, and effort of each requirement, (iii) requirements packaging, that is identifying clusters of requirements, (iv) requirements linking, that is establishing traceability links between requirements and other artifacts, and (v) evaluation, that is utilizing generated traces for certain purposes, e.g. to estimate the impact of change for particular requirements. VB combines a manual and semi-automated way in obtaining the traceability links and in performing a change in the software artifacts. G. Zemont in [11] proposed this
approach.

G. Scenario-Based Approach (SB)

Egyed et al. in [12-14] propose scenario-based approach. SB uses a hypothesized trace information that have to be manually entered. Then, it uses runtime information to creating trace links. Test case scenarios are executed on a running system and execution information is obtained using a monitoring tool. The information is then combined with the hypothesized trace information to form a footprint graph. This graph shows the relationship among artifacts in the system. The traceability links are created automatically, but the hypothesized trace information must be manually entered. In SB, traceability links can only be created once a running system is available.

Researcher’s Work

Based on the above approaches, we are explaining here the work of different researchers related to test coverage measurement and analysis.

Faizah and Ibrahim, 2009 [15] proposed the approach that was specifically designed and implemented to analyze and construct the test coverage. Their framework suggests that Software Requirement Specification (SRS), Software Design Document (SDD) and Software Test Document (STD) are to be used as the sources to get the input for test coverage. These sources of an application are to be instrumented with additional codes as markers. Thus when a test script or a test case is run, echoed markers from the executed application can be collected and processed by a trace analyzer. This analysis will generate the traceability matrices data which is kept in the database for test coverage analysis. The analysis of the test coverage can be done by using the collected data kept in the database.

Sakamoto et al. 2010, [16] proposed a framework for consistent and flexible measurement of test coverage, called the Open Code Coverage Framework (OCCF), which supports multiple programming languages. OCCF extracts commonalities from multiple programming languages focusing on only small syntax differences in programming languages using an abstract syntax tree. It consists of three subsystems: the code insertion, code-execution and coverage-display subsystems. OCCF inserts the measurement code into the source code, and coverage is measured by executing the program. OCCF provides guidelines to support several test coverage criteria.

Kaphammer et al. 2008, [17] presented a test coverage monitoring technique that tracks a program’s definition and use of database entities during test suite execution. The instrumentation phase accepts an adequacy criterion, a test suite, and the program under test. They instrumented the program under test by placing probes at key locations within the Control Flow Graph (CFG). The execution of the instrumented program and test suite yields the database-aware coverage results. The coverage report records the definition and use of relational database entities in the context of both the method and the test case that perform the database interaction.

Kessis et al. 2005, [18] used JOnAS J2EE server as the case study to present test and coverage analysis for large scale applications. They addressed the code coverage, which corresponds to a white box approach where the internal mechanisms of the application under test are seen by the tester. They used Clover coverage analyzer for coverage measurement. The Clover analyzer tasks are fully automated via the Ant tool. The integration task of the coverage measure processes with the build project and test processes are relatively easy. First the JOnAS sources are instrumented. Then a binary version of the instrumented code is generated. Third, tests are run. During test runtime, coverage data are gathered and stored. Finally coverage reports are generated based on stored coverage data.

Lingampally et al., 2007 [19] introduced a Java byte code coverage analyzer which provides test coverage information at method, block, branch and predicate level. The overall system has four components as, i.e. Instrumentor, Database, Code Change Analyzer and Graphical User Interface (GUI). Code under test (CUT) is given to the instrumentation engine, i.e., the Instrumentor, which analyzes the byte code using BCEL APIs and inserts probes at appropriate locations depending on the type of coverage desired. The instrumented code is then run with a test case or a test suite. During a test-execution, when a probe is invoked, the relevant data is recorded in the database.

Lormans and Duersen [20] presented a method for generating requirement coverage views. They studied to investigate to what extent relevant traceability links can be reverse engineered automatically from available documents and used for generating coverage views. They used Latent Semantic Indexing (LSI) for recovering the traceability links. Their focus is on reconstructing coverage views, i.e., views on the requirements set that can be used to monitor the progress in requirements development, design, and testing.

Genetic algorithms work on the basis of ‘genes’, which are created randomly. They are then subjected to some task. Genes demonstrating good performance are kept for next phases, while others are discarded. As far as testing is concerned, Genetic algorithm searches for optimal test parameter combinations that satisfy a predefined test criterion. This test criterion is represented through a “coverage function” that measures how many of the automatically generated optimization parameters satisfy the given test criterion. Genes that optimize the coverage function will survive while others will be discarded; the process is repeated with optimized genes being replicated and further random genes replacing of discarded genes. Ultimately one gene (or a small group of genes) will be left in the set and this would logically be the best fit for coverage function.

Rauf et al., 2010, [21] have proposed a genetic algorithm based technique for coverage analysis of GUI testing. The technique has been subjected to extensive testing. Three simple applications were selected to experiment with, which included Notepad, WordPad, and MS WORD. The results have shown enhanced coverage with an increase in the number of generations.

Ciyong Chen et al., 2009, [22] presented a new method called EPDG-GA which utilizes the Edge Partitions Dominator Graph (EPDG) and Genetic Algorithm (GA) for branch coverage measurement. First, a set of Critical Branches (CBs) are obtained by analyzing the EPDG of the tested program, while covering all the CBs implies covering all the branches of the Control Flow Graph (CFG). Then, the
fitness functions are instrumented in the right position by
analyzing the Pre-Dominator Tree (PreDT), and two metrics
are developed to prioritize the CBs. Coverage-Table is
established to record the CBs information and keeps track of
whether a branch is executed or not. GA is used to generate
test data to cover CBs so as to cover all the branches.

Ricardo D. F. et al. 2010, [23] used UML state machine
model to show visually the coverage achieved by a test suite.
They presented a novel model coverage analysis tool
(MoCAT) that analyses the coverage of a UML state
machine model achieved by a given test suite (either
manually or automatically generated) and represents it
visually. This coverage is represented by colouring the
elements in the model, so it is easier and faster to visualize
the portions of the model that were not exercised by the test
suite.

approach to structural coverage based on a virtualized,
instrumented, execution environment. Their framework,
Couverture, is able to measure structural coverage of object
and source code without requiring any form of application
instrumentation with a single execution of the cross-
compiled application and test suites. The basic idea behind
Couverture is to virtualize the approach commonly used to
measure object code coverage: while traditional object
coverage tools require a physical connection to the target
hardware to measure coverage. Couverture uses a
virtualized environment producing a rich execution trace
containing the address of executed object instructions and
the outcomes of conditional branches. Couverture is then
able to determine actual object code coverage by processing
several execution traces corresponding to the executions of
different tests.

IV. COMPARATIVE EVALUATION

This section provides comparative analysis based on few
important criteria such as coverage types, data base support,
software traceability support, regression testing support,
change management and test case generation support.

<table>
<thead>
<tr>
<th>Researchers</th>
<th>CT</th>
<th>DB Support</th>
<th>ST Support</th>
<th>RT Support</th>
<th>CCT</th>
<th>CRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faizah et al. (2009)</td>
<td>Method, Class,</td>
<td>Yes</td>
<td>Yes Both</td>
<td>No</td>
<td>With and without</td>
<td>Traceability</td>
</tr>
<tr>
<td></td>
<td>Package, Requirement</td>
<td></td>
<td></td>
<td></td>
<td>running a test case</td>
<td>matrix</td>
</tr>
<tr>
<td>Sakamanto et al.(2010)</td>
<td>Statement, Line,</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Condition, Decision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kapfhammer et al.(2008)</td>
<td>Data Flow</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>After running a test</td>
<td>Test re-run</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>case</td>
<td></td>
</tr>
<tr>
<td>Kessis et al.(2005)</td>
<td>Line, Condition</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>After running a test</td>
<td>Test re-run</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>case</td>
<td></td>
</tr>
<tr>
<td>Lingampally et al.(2007)</td>
<td>Block, Branch,</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>With and without</td>
<td>Database coverage</td>
</tr>
<tr>
<td></td>
<td>Method, Class</td>
<td></td>
<td>Code Traceability</td>
<td></td>
<td>running a test case</td>
<td>retrieval</td>
</tr>
<tr>
<td>Lormans et al. (2005)</td>
<td>Requirement</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Traceability links</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Requirement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raul et al. (2010)</td>
<td>Path</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Ciyong Chen et al.(2009)</td>
<td>Branch</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Without running a</td>
<td>Coverage table</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>test case</td>
<td>retrieval</td>
</tr>
<tr>
<td>Ricardo D.F. et al.(2010)</td>
<td>MC/DC</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>After running a test</td>
<td>Test re-run</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Code Traceability</td>
<td></td>
<td>case</td>
<td></td>
</tr>
<tr>
<td>Matteo et al. (2009)</td>
<td>Code coverage</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>After running a test</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Code Traceability</td>
<td></td>
<td>case</td>
<td></td>
</tr>
</tbody>
</table>

NA – Not Available

A. Coverage Types(CT)
The effectiveness of coverage measurement and analysis
depends upon which coverage type has been used by an
approach. The common coverage types used by researchers
include; Statement, Basic Block, Branch, Path, Conditions,
Loop, Condition/Decision, Modified Condition/Decision
(MCDC), Method or Function, Class, Package, Design and
Requirement. Some other coverage types are Data Flow and
Control Flow coverage, c-use and p-use coverage.

B. Data Base(DB) Support
An important aspect of a good approach is its open data
base support. Databases are used to record code elements
and test coverage information for latter retrieval for further
exploration of the facts.

C. Software Traceability (ST) Support
Software traceability is used in order to link software
artifacts (code, requirement, test case) in either ways; ‘code
traceability’ (code to test case) and ‘requirement
traceability’ (requirement to test case). Most approaches use
code to test case to measure test coverage however the use
of both is increasing to measure integrated test coverage.

D. Regression Testing (RT) Support
Regression testing is any type of software testing that
seeks to uncover software errors after changes to the
program (e.g. bug fixes or new functionality) have been
made, by retesting the program. The intent of regression
testing is to assure that a change, such as a bug fix, did not
introduce new bugs [11]. Regression testing is done so as to
execute maximum portion of changed code and achieve
maximum coverage.

E. Coverage Construction Technique (CCT)
Coverage construction is another important criteria for
evaluation. It depends upon whether coverage is constructed
with or without running a test case.

F. Coverage Re-Construction Technique (CRT)
A software change (a change of method, class, package,
design, test cases, requirements, etc.) will lead to changing

This paper aimed to provide general overview and comparison of the current test coverage measurement and analysis approaches. We categorized these approaches into different categories. Then, we have evaluated all these approaches based on some criteria namely, coverage types, data base support, software traceability support, regression testing support, coverage construction and reconstruction techniques (test re-run, traceability link presence calculation, and traceability matrix and database coverage percentage retrieval). Finally, we have summarized important factors based on our study.

V. CONCLUSION

This paper satisfied all the evaluation criteria.