



Metamorphic Testing

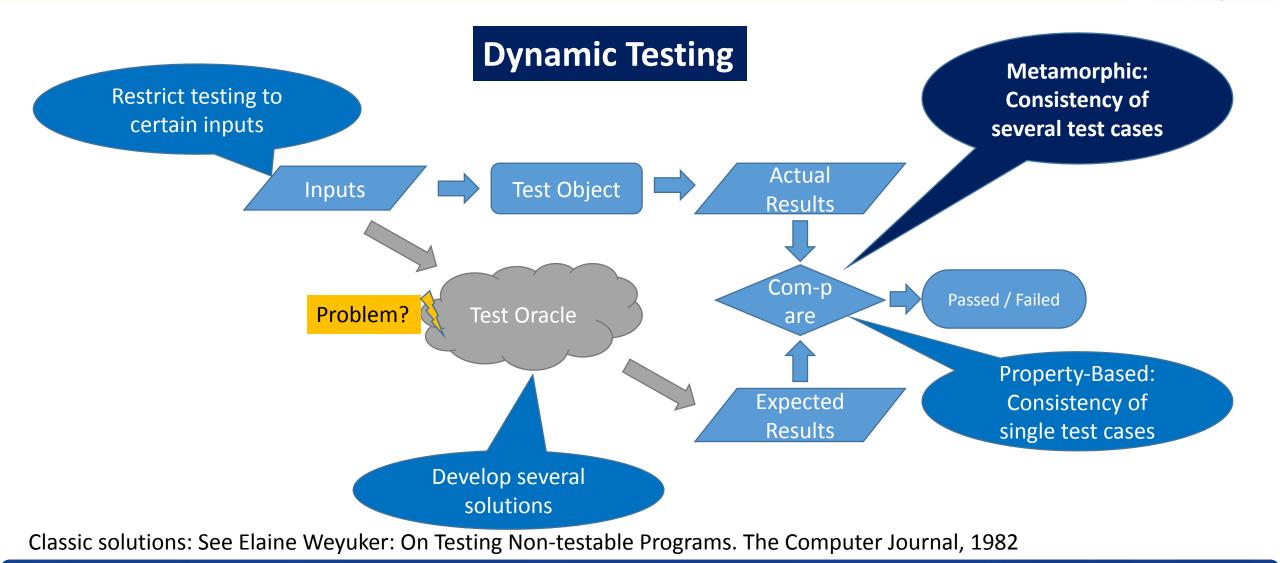
Matthias Hamburg, German Testing Board October 2022



The Oracle Problem in Testing and Solutions

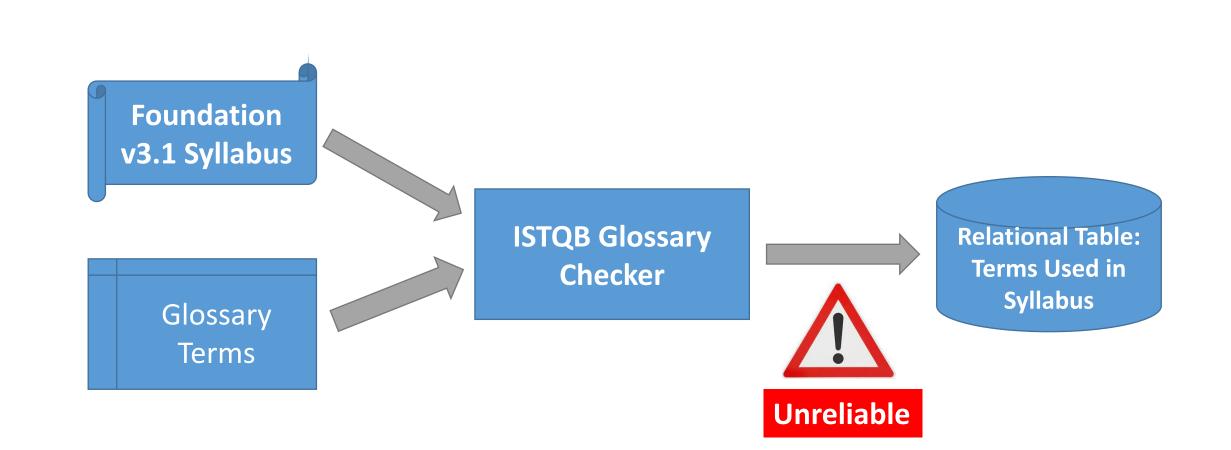


Software. Testing. Excellence.



29.08.2022

Metamorphic Testing From My Own Experience



German Testing Board

Source Test Case TC1: Simple Text Search for "test"

Input:

Simple search for the string "test" in the entire ISTQB Glossary in English

Actual Result: List of 226 terms out of 590 in total

We assume it passed

iSTQB Glossary	× +		- o ×
\leftarrow \rightarrow C \textcircled{a}	O A https://beta.glossary.istqb.org/en/search/test	🗉 🖒 🔍 Suchen	☑ 🕲 🧿 🖤 ≡
ISTOB International Software Testing Qualifications Board	STQB Glossary test	Advanced Options 🗸	Reports English Help
Click on a term to see r	more information about a term		226 Results Found
test	A set of one or more test cases.		
acceptance test-driven o	development A collaborative approach to development ir their requirements, which forms the basis for testing a co		omers own domain language to understand
acceptance testing	A test level that focuses on determining whether to accept	pt the system.	
accessibility testing	Testing to determine the ease by which users with disabi	lities can use a component or system.	
Agile testing	Testing practice for a project using Agile software develo (XP), treating development as the customer of testing an		d methods, such as extreme programming
alpha testing	Simulated or actual operational testing conducted in the	developer's test environment, by roles outside the de	evelopment organization.
analytical test strategy	A test strategy whereby the test team analyzes the test b	asis to identify the test conditions to cover.	
API testing	Testing performed by submitting commands to the softv	vare under test using programming interfaces of the	e application directly.
audio testing	Testing to determine if the game music and sound effect	ts will engage the user in the game and enhance the	game play.

German Testing Board

Metamorphic Relation MR1

~



Software. Testing. Excellence.

MR1: Adding a filter for terms used in a Syllabus must yield a subset of the original result list.

ISTQB Glossary	× +				-	-		×
→ C @	O A https://beta.glossary. istqb.org /en/search/	E ☆ Q Suchen		٥	©	۵	200	=
ISTOB International Software Testing Qualifications Board	ISTQB Glossary test	Q Advanced Options A Repo	orts í	English		Help	þ	
	Check each Syllabus you want to search: Check All / Uncheck All Foundation v3.1 Advanced Test Manager 2012 Advanced Test Analyst 2012 Advanced Test Analyst v3.0 Advanced Test Analyst v3.1 Advanced Technical Test Analyst 2012 Advanced Technical Test Analyst v3.0 Advanced Technical Test Analyst v3.0 Advanced Technical Test Analyst v4.0 Expert Improving the Test Process v1.0.0 Expert Test Management 2011 Foundation Agile Tester 2014 Specialist - Acceptance Testing 2019 Specialist - Automotive Software Tester 2018 Specialist - Mobile Application Testing 2019 Specialist - Model-Based Testing 2015 Specialist - Performance Testing 2018 Specialist - Vasbility Testing 2018 	You must select at least one type: Check All / Uncheck All Terms Acronyms Definitions Synonyms I More Options: Keywords only Non-Keywords only Active Terms only Not-Active Terms only Exact Matches only Multi Language Display (Search in English only) Show English Translation (Search in non-English only) Advanced S	W)					

Follow-up Test Case TC2

GTB German Testing Board

Software. Testing. Excellence.

Input: Same as for TC1 + filter for Foundation v3.1 Syllabus

Actual result: List with 112 terms.

We verify MR1 with a comparator

MR1 Passed.

🛩 ISTQB Glossary	× +						_		×
\leftarrow \rightarrow C \textcircled{a}	O A https://beta.glossary.istqb.org/en/search/test	E 🗘	Q Suchen		${igvarsiments}$	M	C	v	Ξ
ISTOB International Software Testing Qualifications Board	STQB Glossary test	Advan	ced Options 🗸	Reports	En	glish		Help	
Click on a term to see n	nore information about a term			_			112 Re	sults Found	
test	A set of one or more test cases.						11210	Suito i Guin	-
acceptance testing	A test level that focuses on determining whether to accept the system.								
accessibility testing	Testing to determine the ease by which users with disabilities can use a comp	ponent or	r system.						
alpha testing	Simulated or actual operational testing conducted in the developer's test envi	ronment,	by roles outside the o	development o	rganizat	ion.			
beta testing	A type of acceptance testing performed at an external site to the developer's	test envir	onment by roles outs	ide the develop	oment o	rganiza	ation.		
black-box test technique	A test technique based on an analysis of the specification of a component of	o <mark>r system</mark>	า.						
change-related testing	A type of testing initiated by modification to a component or system.								
checklist-based testing	An experience-based test technique whereby the experienced tester uses a h rules or criteria against which a product has to be verified.	igh-level	list of items to be not	ed, checked, o	r remem	bered,	oras	et of	
component integration to	esting Testing in which the test items are interfaces and interactions betwee	n integra	ited components.						

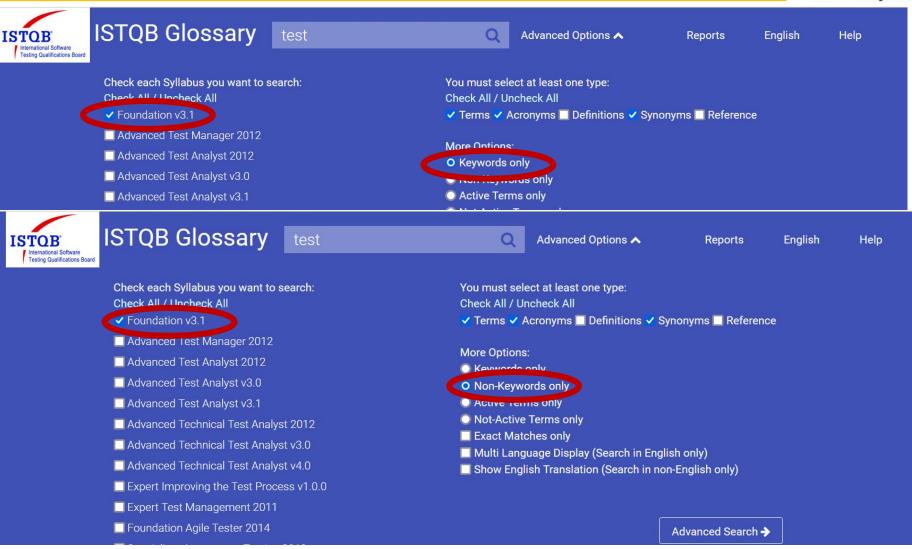
Metamorphic Relation MR2



Software. Testing. Excellence.

Let us make TC2 to a source test case of a new metamorphic relation.

MR2: Adding the filters for Keywords only/Non-Keywords only must partition the output of the source test case.



© GTB

Follow-up Test Cases TC3 and TC4



Software. Testing. Excellence.

64 Results Found

Click on a term to see more information about a term

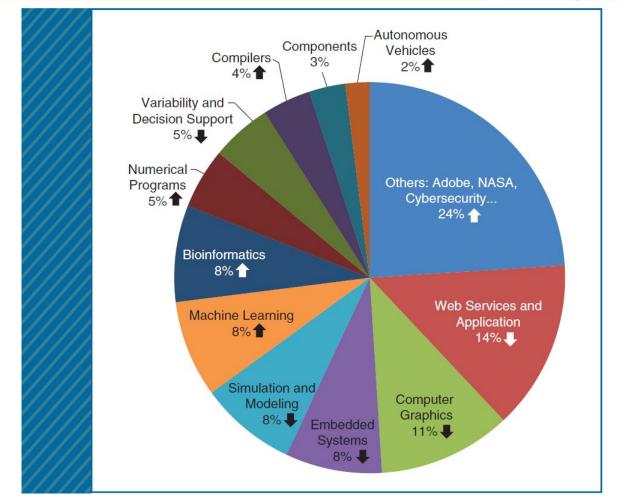
Input:	acceptance testing	A test level that focuses on determining whether to accept the system.
Same as for TC2 +	alpha testing	Simulated or actual operational testing conducted in the developer's test environment, by roles outside the development organization.
filters for Keywords	beta testing	A type of acceptance testing performed at an external site to the developer's test environment by roles outside the development organization.
only / Non-keywords	black-box test technique	A test technique based on an analysis of the specification of a component or system.
only	change-related testing	A type of testing initiated by modification to a component or system.
Actual results:	checklist-based testing	An experience-based test technique whereby the experienced tester uses a high-level list of items to be noted, checked, or remembered, or a set of rules or criteria against which a product has to be verified.
Keywords only: 64 Non-keywords only: 42	Click on a term to see m	nore information about a term 42 Results Found
	test	A set of one or more test cases.
We verify MR2:	accessibility testing	Testing to determine the ease by which users with disabilities can use a component or system.
Expected: 112 in total.	decision testing	A white-box test technique in which test cases are designed to execute decision outcomes.
MR2 failed:	exhaustive testing	A test approach in which the test suite comprises all combinations of input values and preconditions.
6 terms are missing!	experience-based testing	Testing based on the tester's experience, knowledge and intuition.
	high-level test case	A test case with abstract preconditions, input data, expected results, postconditions, and actions (where applicable).

Publications About Metamorphic Testing



Software. Testing. Excellence

- Introduced in 1998 by Chen et al.
- Survey conducted in 2018: 84 case studies published
- Included in ISO/IEC/IEEE 29119-4:2021 (Test Techniques)
- Not discussed (yet) in ISTQB Advanced Test Analyst



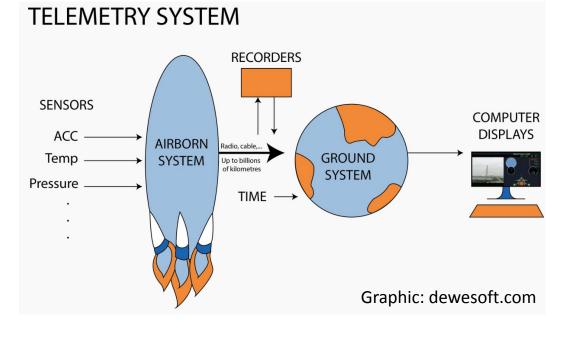
Distribution of 84 case studies by application domains Source: Segura et al., 2020

Lindvall et al., 2015: Accuracy of the NASA Data Access Toolkit

Software. Testing. Excellence.

German Testing Board

- Test object: NASA Data Access Toolkit
 - High-performance query and research user interface
 - Huge database of telemetry data of the NASA missions
- Oracle problem:
 - Are the query results accurate?
- Metamorphic relations:
 - Equivalent queries must deliver the same outputs
- Test result:
 - A large number of defects detected
 - Successfully combined with MBT and test automation in an agile software lifecycle



Zhou et al., 2016: Reliability of Search Engines

• Test objects: The most popular search engines • Google, Bing, Chinese Bing and Baidu The database is huge and changes all the time • Metamorphic relations: Restriction to the url domains in the result list • Restriction to the titles in the result list Permutations of conditions combined with "AND"

• Etc.

• Oracle problem:

- Test result:
 - None of the search engines is 100% reliable
 - In Englisch is Google the most reliable, in Chinese it is Baidu
 - The testers reported several specific defects to the developers

Web	Images	Videos	News	Shopping	More *	Search tools
About 8	results (0.28	seconds)				
books.g Joseph	oogle.com.a Story - 1805	u/books?id - Civil proc mpted pea	edure	t he the Plaint	iff, at the hi 1	time w Vn, &c, a
Indian news.go tempted	ogle.com/ne d peaceably	orrespore wspapers'	ndence. 2nid=23268 ut force to d	- Google I adat=1887030	Vews 4&id titution, whic	h de . clares that I
Indiar news.go tempted shall act	apolis C oogle.com/nd d peaceably t as Presider	orrespor wspapers' and withont, of the S	ndence. 2nid=23268 ut force to o enate. But	- Google I adat=1887030 obey the Cons he was resiste (a)	Vews 4&id titution, whic	h de . clares that I
Indiar news.go tempted shall act	apolis C ogle.com/ne d peaceably	orrespor wspapers' and withont, of the S	ndence. 2nid=23268 ut force to o enate. But	- Google I adat=1887030 obey the Cons he was resiste (a)	Vews 4&id titution, whic	h de . clares that I

About 44,000 results (0.37 seconds)

No results found for "tempted peaceably" site:au.

Results for tempted peaceably site:au (without quotes):

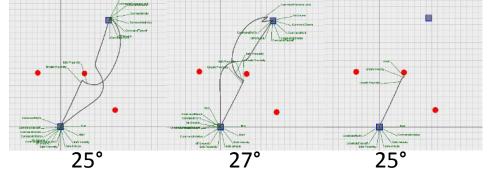
Oracle problems of AI-based Systems

- The expected result is often unspecified
 - The behavior cannot be predicted exactly
 - The behavior may change over time without changing the software
- Metamorphic testing is a test technique of choice
 - Dealt with by the new ISTQB[®] Syllabus "Certified Tester AI Testing"
 - Included in the standard ISO/IEC TR 29119-11:2020 (Testing of AI-based Systems)



Lindvall et al., 2017: Safety of Autonomous Drones

- Test object: Drone AI controller
 - Tested in a simulated environment
- Oracle problem:
 - Does the autonomous drone behave safely and as expected?
- Metamorphic relations:
 - Variations of the exposure to light; Rotation; Translation; Changing the distance of the obstacles; Geometrically equivalent obstacle groups
- Test result:
 - Failures in certain situations: Collision with obstacles, failing to recognize the landing pad, unstable behavior, etc.



Follow-up test cases rotating the scenery by three similar angles. Source: Lindvall et al.

German Testing Board

Tian et al., 2018: Correctness of Autonomous Driving



Software.Testing.Excellence.

- Test objects: Autonomous vehicle control systems
 - Three top-performing deep neural networks of the Udacity self-driving car challenge
- Oracle problem:
 - What is the correct driving behavior in a specific situation?
- Metamorphic relations:
 - Same driving task under various environmental and camera conditions: rain, fog, blurring, picture scale, contrast, brightness, rotation, translation, shear, etc.
- Test result:
 - Hundreds of failures that might lead to severe or even fatal accidents
 - Neuron coverage is relevant for effective defect detection



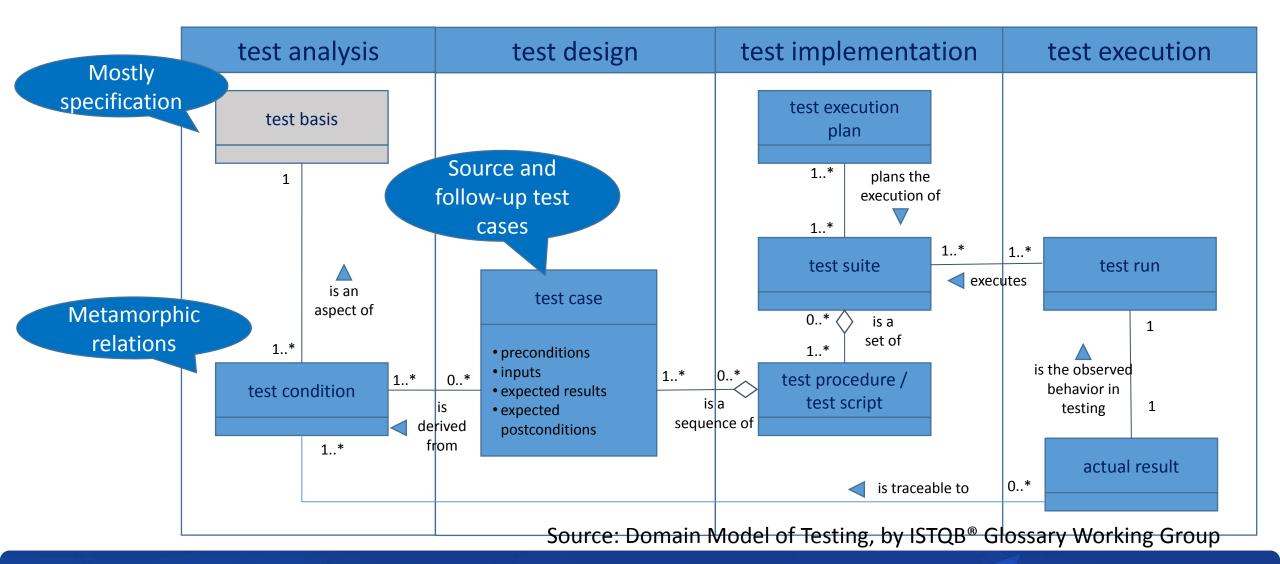
Source test case

Rain added

Source: deeplearningtest.github.io/deepTest/

Metamorphic Testing Is a Test Technique





Analyzing Metamorphic Relations...

- ... determines the effectiveness of defect detection
- ... is an essential creative activity
 - Knowledge of the application domain
 - Functional decomposition
 - Formal specification of metamorphic relations
 - Highest possible code coverage
- ... has two different approaches
 - Input-oriented: Which change of the input leads to a verifiable change of the output?
 - Output-oriented: When are the expected outputs identical / equivalent / partitioning etc.?







- ISO/IEC/IEEE 29119-4 (2021):
 - "There is currently no industry agreed approach to calculating coverage for metamorphic testing."
- We need alternatives to the coverage-based end criteria
 - Option 1: Defect detection rate (e.g. number of failures per 100 follow-up test cases)
 - Option 2: White-box coverage (e.g. neuron coverage)



- A relatively new test technique
- Especially suitable for test objects with a relevant oracle problem
 - Web services, artificial intelligence
- Integrates well with MBT, Agile and DevOps
- The defect detection effectiveness is empirically proven
- The efficiency depends on the degree of automation

Does metamorphic testing deserve to be included in the canon of test techniques?



Thank you very much for your attention!

German Testing Board

President Florian Fieber

Vice-Presidents Horst Pohlmann Dietrich Leimsner

Executive Director Dr. Armin Metzger

Back Office

Andrea Kränzlein OFFICE-MANAGEMENT Koldestrasse 8 b 91052 Erlangen GERMANY Tel.: +49 (0) 91 31 / 97 61 06 Mobile: +49 (0) 171 / 63 90 749 E-Mail: backoffice@qtb.de

Matthias Hamburg



