CONTINUOUS INTEGRATION IN GITHUB QUALITY AND PRODUCTIVITY OUTCOMES



Bogdan Vasilescu



Yue Yu



Prem Devanbu



Vladimir Filkov





SOFTWARE PROCESSES ARE SOFTWARE TOO

Leon Osterweil

University of Colorado Boulder, Colorado USA

2

1. The Nature of Process.

The major theme of this meeting is the exploration of the importance of .ul process as a vehicle for improving both the quality of software products and the the way in which we develop and evolve them. In beginning this exploration it seems important to spend at least a short time examining the nature of process and convincing ourselves that this is indeed a promising vehicle.

We shall take as our elementary notion of a process that it is a systematic approach to the creation of a product or the accomplishment of some task. We observe that this characterization describes the notion of process commonly used in operating systems-- namely that a process is a computational task executing on a single computing device. Our characterization is much broader, however, describing any mechanism used to carry out work or achieve a goal in an orderly way. Our processes need not even be executable on a computer.

It is important for us to recognize that the notion of process is a pervasive one in the realm of human activities and that humans seem particularly adept at creating and carrying out processes. Knuth [Knuth 69] has observed that following recipes for food preparation is an example of carrying out what we now characterize as a process. Similarly it is not difficult to see that following assembly instructions in building toys or modular furniture is carrying out a process. Following office procedures or pursuing the steps of a manufacturing activity are more widely understood to be the pursuit of orderly process.

The latter examples serve to illustrate an important pointnamely that there is a key difference between a process and a process description. While a process is a vehicle for doing a job, a process description is a specification of how the job is to be done. Thus cookbook recipes are process descriptions while the carrying out of the recipes are processes. Office procedure manuals are process descriptions, while getting a specific office task done is a process. Similarly instructions for how to drive from one location to another are process descriptions, while doing the actual navigation and piloting is a process. From the point of view of a computer scientist the difference can be seen to be the difference between a type or class and an instance of that type or class. The process

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission. description defines a class or set of objects related to each other by virtue of the fact that they are all activities which follow the dictated behavior. We shall have reason to return to this point later in this presentation.

For now we should return to our consideration of the intuitive notion of process and study the important ramifications of the observations that 1) this notion is widespread and 2) exploitation of it is done very effectively by humans. Processes are used to effect generalized, indirect problem solving. The essence of the process exploitation paradigm seems to be that humans solve problems by creating process descriptions and then instantiating processes to solve individual problems. Rather than repetitively and directly solving individual instances of problems, humans prefer to create generalized solution specifications and make them available for instantiation (often by others) to solve individual problems directly.

One significant danger in this approach is that the process itself is a dynamic entity and the process description is a static entity. Further, the static process description is often constructed so as to specify a very wide and diverse collection of dynamic processes. This leaves open the distinct possibility that the process description may allow for process instances which do not perform "correctly." Dijkstra makes this observation in his famous letter on the GOTO statement, [Dijkstra 69] observing that computer programs are static entities and are thus easier for human minds to comprehend, while program executions are dynamic and far harder to comprehend and reason about effectively. Dijkstra's point was important then and no less significant now. Processes are hard to comprehend and reason about, while process descriptions, as static objects, are far easier to comprehend. Finally it is important to also endorse Dijkstra's conclusion that our reasoning about process descriptions is increasingly useful in understanding processes as the descriptions are increasingly transparent descriptions of all processes which might be instantiated.

In view of all of these dangers and difficulties it is surprising that humans embark upon the indirect process description/instantiation/execution approach to problem solving so frequently. It is even more startling to observe that this approach is successful and effective so often. This suggests that humans have an innate facility for indirect problem solving through process specification. It is precisely this innate ability which should be able to propel us to become far more systematic and effective in the development and evolution of computer software. What currently stands most directly in our way is our failure--to date--to understand our most central and difficult problems in terms of the process description/instantiation/execution paradigm.

ICSE, 1987



© 1987 ACM 0270-5257/87/0300/0002\$00.75

SOFTWARE PROCESSES ARE SOFTWARE TOO

Leon Osterweil

University of Colorado Boulder, Colorado USA

2

1. The Nature of Process.

The major theme of this meeting is the exploration of the importance of .ul process as a vehicle for improving both the quality of software products and the the way in which we develop and evolve them. In beginning this exploration it seems important to spend at least a short time examining the nature of process and convincing ourselves that this is indeed a promising vehicle.

We shall take as our elementary notion of a process that it is a systematic approach to the creation of a product or the accomplishment of some task. We observe that this characterization describes the notion of process commonly used in operating systems-- namely that a process is a computational task executing on a single computing device. Our characterization is much broader, however, describing any mechanism used to carry out work or achieve a goal in an orderly way. Our processes need not even be executable on a computer.

It is important for us to recognize that the notion of process is a pervasive one in the realm of human activities and that humans seem particularly adept at creating and carrying out processes. Knuth [Knuth 69] has observed that following recipes for food preparation is an example of carrying out what we now characterize as a process. Similarly it is not difficult to see that following assembly instructions in building toys or modular furniture is carrying out a process. Following office procedures or pursuing the steps of a manufacturing activity are more widely understood to be the pursuit of orderly process.

The latter examples serve to illustrate an important pointnamely that there is a key difference between a process and a process description. While a process is a vehicle for doing a job, a process description is a specification of how the job is to be done. Thus cookbook recipes are process descriptions while the carrying out of the recipes are processes. Office procedure manuals are process descriptions, while getting a specific office task done is a process. Similarly instructions for how to drive from one location to another are process descriptions, while doing the actual navigation and piloting is a process. From the point of view of a computer scientist the difference can be seen to be the difference between a type or class and an instance of that type or class. The process

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission. description defines a class or set of objects related to each other by virtue of the fact that they are all activities which follow the dictated behavior. We shall have reason to return to this point later in this presentation.

For now we should return to our consideration of the intuitive notion of process and study the important ramifications of the observations that 1) this notion is widespread and 2) exploitation of it is done very effectively by humans. Processes are used to effect generalized, indirect problem solving. The essence of the process exploitation paradigm seems to be that humans solve problems by creating process descriptions and then instantiating processes to solve individual problems. Rather than repetitively and directly solving individual instances of problems, humans prefer to create generalized solution specifications and make them available for instantiation (often by others) to solve individual problems directly.

One significant danger in this approach is that the process itself is a dynamic entity and the process description is a static entity. Further, the static process description is often constructed so as to specify a very wide and diverse collection of dynamic processes. This leaves open the distinct possibility that the process description may allow for process instances which do not perform "correctly." Dijkstra makes this observation in his famous letter on the GOTO statement, [Dijkstra 69] observing that computer programs are static entities and are thus easier for human minds to comprehend, while program executions are dynamic and far harder to comprehend and reason about effectively. Dijkstra's point was important then and no less significant now. Processes are hard to comprehend and reason about, while process descriptions, as static objects, are far easier to comprehend. Finally it is important to also endorse Dijkstra's conclusion that our reasoning about process descriptions is increasingly useful in understanding processes as the descriptions are increasingly transparent descriptions of all processes which might be instantiated.

In view of all of these dangers and difficulties it is surprising that humans embark upon the indirect process description/instantiation/execution approach to problem solving so frequently. It is even more startling to observe that this approach is successful and effective so often. This suggests that humans have an innate facility for indirect problem solving through process specification. It is precisely this innate ability which should be able to propel us to become far more systematic and effective in the development and evolution of computer software. What currently stands most directly in our way is our failure--to date--to understand our most central and difficult problems in terms of the process description/instantiation/execution paradigm.

ICSE, 1987



© 1987 ACM 0270-5257/87/0300/0002\$00.75

SOFTWARE PROCESSES ARE SOFTWARE TOO

Leon Osterweil

University of Colorado Boulder, Colorado USA

2

1. The Nature of Process.

The major theme of this meeting is the exploration of the importance of .ul process as a vehicle for improving both the quality of software products and the the way in which we develop and evolve them. In beginning this exploration it seems important to spend at least a short time examining the nature of process and convincing ourselves that this is indeed a promising vehicle.

We shall take as our elementary notion of a process that it is a systematic approach to the creation of a product or the accomplishment of some task. We observe that this characterization describes the notion of process commonly used in operating systems-- namely that a process is a computational task executing on a single computing device. Our characterization is much broader, however, describing any mechanism used to carry out work or achieve a goal in an orderly way. Our processes need not even be executable on a computer.

It is important for us to recognize that the notion of process is a pervasive one in the realm of human activities and that humans seem particularly adept at creating and carrying out processes. Knuth [Knuth 69] has observed that following recipes for food preparation is an example of carrying out what we now characterize as a process. Similarly it is not difficult to see that following assembly instructions in building toys or modular furniture is carrying out a process. Following office procedures or pursuing the steps of a manufacturing activity are more widely understood to be the pursuit of orderly process.

The latter examples serve to illustrate an important pointnamely that there is a key difference between a process and a process description. While a process is a vehicle for doing a job, a process description is a specification of how the job is to be done. Thus cookbook recipes are process descriptions while the carrying out of the recipes are processes. Office procedure manuals are process descriptions, while getting a specific office task done is a process. Similarly instructions for how to drive from one location to another are process descriptions, while doing the actual navigation and piloting is a process. From the point of view of a computer scientist the difference can be seen to be the difference between a type or class and an instance of that type or class. The process

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission. description defines a class or set of objects related to each other by virtue of the fact that they are all activities which follow the dictated behavior. We shall have reason to return to this point later in this presentation.

For now we should return to our consideration of the intuitive notion of process and study the important ramifications of the observations that 1) this notion is widespread and 2) exploitation of it is done very effectively by humans. Processes are used to effect generalized, indirect problem solving. The essence of the process exploitation paradigm seems to be that humans solve problems by creating process descriptions and then instantiating processes to solve individual problems. Rather than repetitively and directly solving individual instances of problems, humans prefer to create generalized solution specifications and make them available for instantiation (often by others) to solve individual problems directly.

One significant danger in this approach is that the process itself is a dynamic entity and the process description is a static entity. Further, the static process description is often constructed so as to specify a very wide and diverse collection of dynamic processes. This leaves open the distinct possibility that the process description may allow for process instances which do not perform "correctly." Dijkstra makes this observation in his famous letter on the GOTO statement, [Dijkstra 69] observing that computer programs are static entities and are thus easier for human minds to comprehend, while program executions are dynamic and far harder to comprehend and reason about effectively. Dijkstra's point was important then and no less significant now. Processes are hard to comprehend and reason about, while process descriptions, as static objects, are far easier to comprehend. Finally it is important to also endorse Dijkstra's conclusion that our reasoning about process descriptions is increasingly useful in understanding processes as the descriptions are increasingly transparent descriptions of all processes which might be instantiated.

In view of all of these dangers and difficulties it is surprising that humans embark upon the indirect process description/instantiation/execution approach to problem solving so frequently. It is even more startling to observe that this approach is successful and effective so often. This suggests that humans have an innate facility for indirect problem solving through process specification. It is precisely this innate ability which should be able to propel us to become far more systematic and effective in the development and evolution of computer software. What currently stands most directly in our way is our failure--to date--to understand our most central and difficult problems in terms of the process description/instantiation/execution paradigm.



Productivity?

Quality?

© 1987 ACM 0270-5257/87/0300/0002\$00.75





can bush • . •

... traditionally





Create a branch



Add commits



Open a Pull Request



Discussion & Code review



Pull Request updates



Merge

... modernly



... modernly

le review



E Kalliamvakou, D Damian, K Blincoe, L Singer, DM German. ICSE 2015

• Work practices and challenges in pull-based development: the integrator's perspective G Gousios, A Zaidman, MA Storey, A Van Deursen. *ICSE 2015*

Considerable review load

rails / rails	Watch	Star 26,093 % For	k 10,339
Issues Pull requests Labels Milestones Filters - Q is:pr is:open		New pull request	<>
እ 467 Open ✓ 12,551 Closed Author -	Labels Milestones	Assignee ▼ Sort ▼	()
Move Integer#positive? and Integer#negative? query methods to Numeric #20143 opened an hour ago by meinac		P 2	ກ
Deprecate `assert_template`. #20138 opened 9 hours ago by tgxworld		P 8	4~
Add Enumerable#map_with to ActiveSupport #20134 opened 13 hours ago by mlarraz			<u>E II II I</u>
Allow creating a save callback for same name with parent association #20127 opened 23 hours ago by meinac		2	
ActiveSupport::HashWithIndifferentAccess select and reject should return en #20125 opened a day ago by imanel	umerator if called without I	block	
Don't ignore false values for `include_blank` passed to `Tags::Base#select_co #20124 opened a day ago by greysteil	ontent_tag` 🗸	9	
 Fix for irregular inflection inconsistency #20123 opened a day ago by yoongkang 			
Add openssl_verify_mode and sync other smtp_settings with API docs #20117 opened 2 days ago by jfine			

Considerable review load



 Wait for it: Determinants of pull request evaluation latency on GitHub Y Yu, H Wang, V Filkov, P Devanbu, B Vasilescu. *MSR 2015*





Pull Request is automatically merged into testing branch



Test suite runs automatically



Pull Request is updated in response to test failures



Tests rerun after update







Merge

Merge after CI tests pass

GitHub This repository Search	Explore Featu	ures Enterprise	Pricing	Sign	n up	Sign in
rails / rails		• Watch	2,003	star 27,550	[%] Fork	11,060
Issues Pull requests Labels Milestones	Q is:pr is:closed is:	merged		New pull r	equest	$\langle \rangle$
X Clear current search query, filters, and sorts						(!)
Ĵ 8,842 Total	Author -	Labels - Mil	estones 🔻 🛛 A	Assignee 🔻	Sort 🔻	ຳ
removing unecessary default parameter in private method #18356 opened on Jan 6 by georgemillo	~					4-
Documenting 'remove_possible_method' and 'redefine_method' and	ethod' [ci skip]					<u>.lı</u>
Improve protect_from_forgery documentation. #18354 opened on Jan 6 by simi						
Propagate bind_values from join in subquery #18350 opened on Jan 5 by brainopia					₽ 5	
Fix rollback of primarykey-less tables #18349 opened on Jan 5 by jdelStrother					Ç 9	
Switching SecureTokens to Base58 #18347 opened on Jan 5 by robertomiranda					ÇJ 17	
Fix TypeError in Fixture creation #18345 opened on Jan 5 by mtthgn						
Clean up secure_token_test #18344 opened on Jan 5 by jonatack					, □ 3	

Merge after CI tests pass Code review



 Wait for it: Determinants of pull request evaluation latency on GitHub Y Yu, H Wang, V Filkov, P Devanbu, B Vasilescu. MSR 2015

Merge after CI tests pass Code review



 Wait for it: Determinants of pull request evaluation latency on GitHub Y Yu, H Wang, V Filkov, P Devanbu, B Vasilescu. MSR 2015



Clas gatekeeper

"[cl] enables us to automate more of our process which frees us up to focus on the important things — like implementing and shipping features! [...]

[The integration of Travis-cl in GitHub] enables the team to rapidly find integration errors or regression failures in the test suite. This tightens the feedback loop and not only enables more defect free code, but greatly speeds up our process."

Automate All The Things: Continuous Integration and Deployment at RevUnit
 Addam Hardy. <u>http://addamhardy.com/blog/2013/09/28/automate-all-the-things-continuous-integration-and-continuous-deployment-at-revunit/</u>

(2) How does Cl affect software quality?

20% more pull requests merged & 40% fewer rejected (2) How does Cl affect software quality?

50% more bugs reported monthly by <u>core dev's</u>

No impact on bugs reported by <u>externals</u>



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-Cl





- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-CI

Data Set



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-CI

Data Set



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-CI



Data Set



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-CI



Data Set



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-CI



Data Set



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-CI

Data Set



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++

- 200+ pull requests
- Travis-Cl



Data Set



- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-Cl







- 200+ pull requests
- Travis-Cl





- 200+ pull requests
- Travis-CI

Pull Requests



Pull Requests









Pull

Requests

Merged

Rejected





 Influence of social and technical factors for evaluating contribution in GitHub J Tsay, L Dabbish, J Herbsleb. *ICSE 2014*





 Influence of social and technical factors for evaluating contribution in GitHub J Tsay, L Dabbish, J Herbsleb. *ICSE 2014*



Models



Models





Results



Results

Results

More pull requests merged & fewer rejected

(2) How does Cl affect software quality?

More pull requests merged & fewer rejected

Data Set

- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-Cl
- 100+ issues reported (75%+ tagged)

Data Set

- Not forks
- Ruby, Python, JavaScript, PHP, Java, Scala, C, C++
- 200+ pull requests
- Travis-CI
- 100+ issues reported (75%+ tagged)

"Bug" reports

 STM32L1 get_cpuid() hard faults when using a Cat. 1 or
 New issue

 Cat. 2 STM32L1 #3692

 Image: Closed DipSwitch opened this issue 12 days ago · 2 comments

FREE BRADLEY	DipSwitch commented 12 days ago	Labels
	From the STM32L1 Reference Manual (31.2 Unique device ID registers (96 bits)):	arm bug
	Base address: 0x1FF80050 for Cat.1 and Cat.2 devices and 0x1FF800D0 for Cat.3, Cat.4, Cat.5 and Ca	Milestone
	Three solutions possible for this problem:	Release 2015.09
	 Compile time: Via the linkerscript for the device (this I would prefer since this is the cleanest solution in my opinion) 	Assignee
	<pre>MEMORY { rom (rx) : ORIGIN = 0x08000000, LENGTH = 128K ram (rw) : ORIGIN = 0x20000000, LENGTH = 32K cpuid (r) : ORIGIN = 0x1FF80050, LENGTH = 12 }</pre>	Notifications Subscribe You're not receiving notifications from this thread.
	_cpuid_address = ORIGIN(cpuid); INCLUDE cortexm_base.ld	4 participants

From insiders

From insiders

From outsiders

Models

"Bug" reports/month

STM Cat.	32L1 get_cpuid() hard faults when using a Cat. 1 or 2 STM32L1 #3692	New issue
Close	DipSwitch opened this issue 12 days ago · 2 comments	
ff.	DipSwitch commented 12 days ago	Labels
EEBRACLEY	From the STM32L1 Reference Manual (31.2 Unique device ID registers (96 bits)):	arm bug
	Base address: 0x1FF80050 for Cat.1 and Cat.2 devices and 0x1FF800D0 for Cat.3, Cat.4, Cat.5 and Ca	Milestone
	Three solutions possible for this problem:	Release 2015.09
	Compile time: Via the linkerscript for the device (this I would prefer since this is the cleanest solution in my opinion)	Assignee
	<pre>MEMORY { rom (rx) : ORIGIN = 0x08000000, LENGTH = 128K ram (rw) : ORIGIN = 0x20000000, LENGTH = 32K cpuid (r) : ORIGIN = 0x1FF00050, LENGTH = 12 } _cpuid_address = ORIGIN(cpuid); INCLUDE cortexm_base.ld</pre>	Notifications Image: Notifications You're not receiving notifications from this thread. 4 participants

From insiders

+48%

From outsiders

"Bug" reports/month

STM Cat.	32L1 get_cpuid() hard faults when using a Cat. 1 or 2 STM32L1 #3692	New issue
() Close	d DipSwitch opened this issue 12 days ago · 2 comments	
	DipSwitch commented 12 days ago	Labels
FREE BRADLEY	From the STM32L1 Reference Manual (31.2 Unique device ID registers (96 bits)):	arm bug
	Base address: 0x1FF80050 for Cat.1 and Cat.2 devices and 0x1FF800D0 for Cat.3, Cat.4, Cat.5 and Ca	Milestone
	Three solutions possible for this problem:	Release 2015.09
	Compile time: Via the linkerscript for the device (this I would prefer since this is the cleanest solution in my opinion)	Assignee
	MEMORY { rom (rx) : ORIGIN = 0x88000000, LENGTH = 128K	Notifications
	<pre>ram (rw) : ORIGIN = 0x200000000, LENGTH = 32K cpuid (r) : ORIGIN = 0x1FF80050, LENGTH = 12 }</pre>	You're not receiving notifications from this thread.
	_cpuid_address = ORIGIN(cpuid);	4 participants
	INCLUDE cortexm_base.ld	2 2 2

CONTINUOUS INTEGRATION IN GITHUB **QUALITY AND PRODUCTIVITY OUTCOMES**

More pull requests merged & fewer rejected

Travis-Cl (2) How does CI **Results** affect software quality? From insiders From outsiders "Bug" reports/month +48% Notifications 41 Subscribe You're not monthly actifications from the t performante Statistica More bugs reported No impact on bugs monthly by core dev's reported by externals

Bogdan Vasilescu

Vladimir

Filkov

