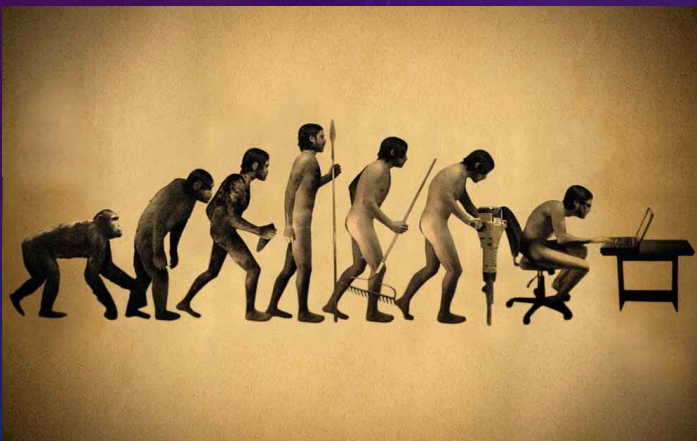


Finding and Fixing Design Debt

Rick Kazman
University of Hawaii

1

My Personal Evolution



SAAM (1994)
ATAM (1998)
CBAM (2001)

2

My Personal Evolution



When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.

- William Thompson, Lord Kelvin

3

My Personal Evolution



There is no sense in being precise when you don't even know what you're talking about.

--John von Neumann

4

My Focus Today



- Automated SE
- Empirical studies
- Reproducibility
- Industrial relevance
- Usability

In theory, there is no difference between theory and practice. In practice, there is. - Yogi Berra

5

My Focus Today

Finding and Fixing Design Debt

6

The State of the Practice



7

The State of the Practice

- The boat is leaking but you keep paddling!
- Why?
 - The illusion of progress.
 - The lack of measurements.
 - Design is largely invisible.



8

Architecture/Design Flaws



9

My "Grand Research Challenge"

- How to measure the health of an architecture?
- Can this be:
 - Automated?
 - Empirically justified?
 - Repeatable?

10

Isn't This a Solved Problem?

- Just use existing TD detection tools, e.g.



11

Sadly, no...

- Results of a recent study:
 - TD detection tools disagree about basic (seemingly) objective measures due to different definitions of fundamental concepts.
 - The majority of what is reported by these tools is no more insightful than LOC.

[Lefever et al. ICSE 2021]

12

There is Hope

- Introducing



<https://archdia.com/>

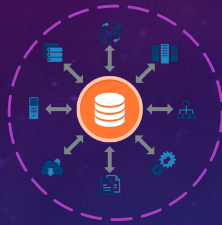
13

Empirical Basis

- ✓ >300 Open Source Projects
- ✓ >50 Industrial projects

14

DV8 Work Flow



Step 1: Data Collection

Code dependency, history,
issue records



Step 2: Automated Analysis

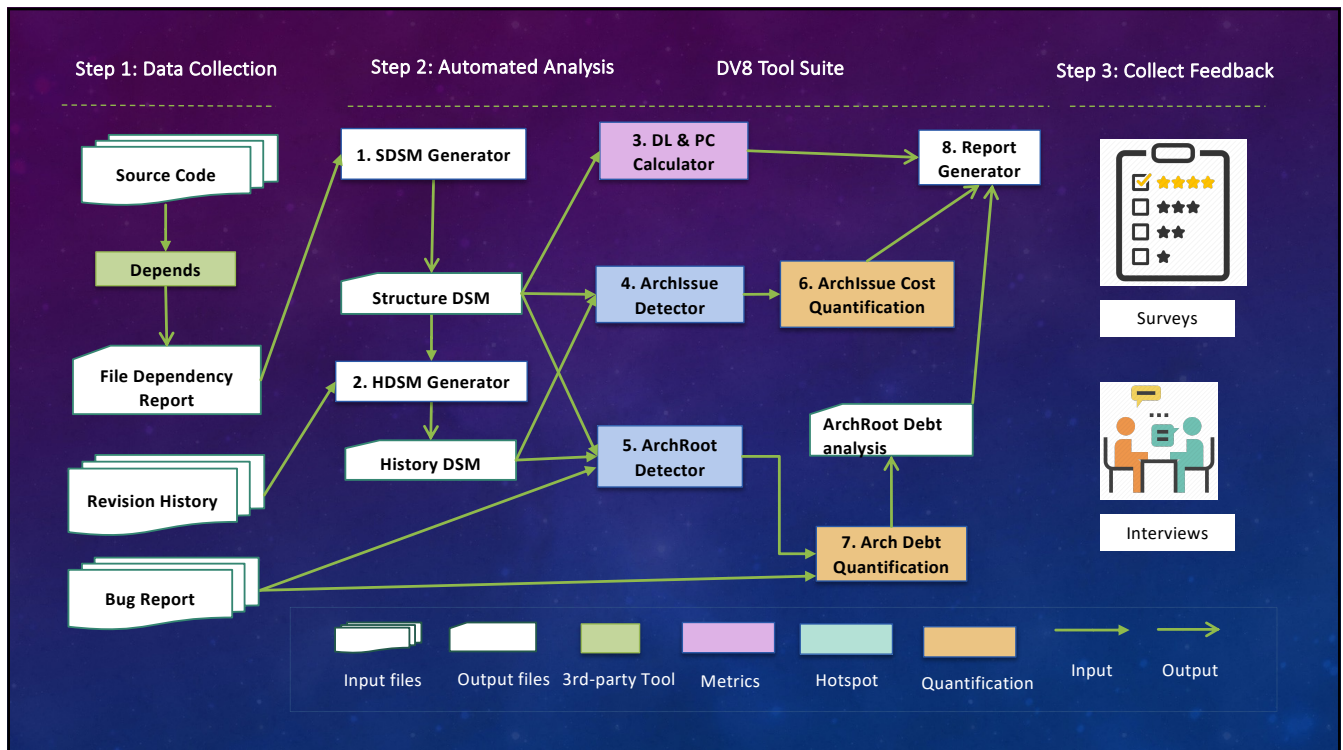
Measurement, hotspot
detection, cost calculation



Step: Collect Feedback

Surveys and Interviews
with practitioners

15



16

Step 1: Data Collection



- ✓ Dependency information
- ✓ History information
- ✓ Issue information

17

Design Rule Space (DRSpace)

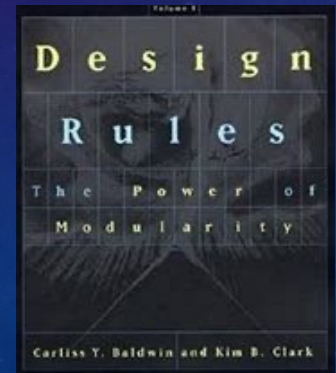


A DRSpace is composed of a **meaningful subset** of a system's files and the **architectural connections** among these files.

- Any subset of files may form a design space
- Architectural connections
 - Structural couplings: call, inherit, aggregate, etc.
 - Evolutionary couplings
 - Implicit or explicit

[Xiao et al, TSE 2018]

[Xiao et al, ICSE 2014]



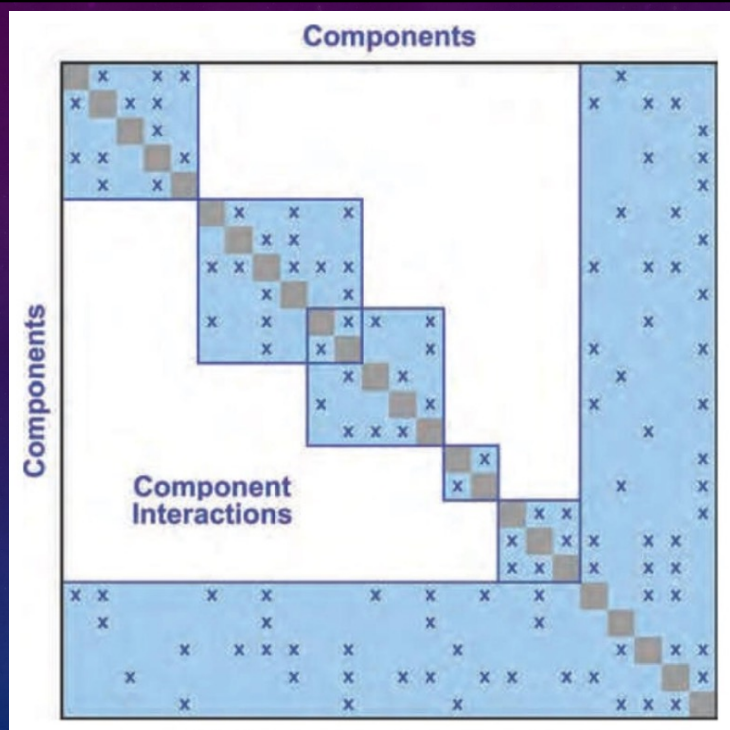
18

Design Rule Space (DRSpace)

- ▶ Non-trivial software system contain multiple design spaces:
 - each feature implemented
 - each pattern applied
 - each concern addressed
- ▶ Each file can participate in multiple DRSpaces
- ▶ Architectures can be modeled as overlapping DRSpaces
- ▶ We visualize each DRSpace as a **Design Structure Matrix (DSM)**

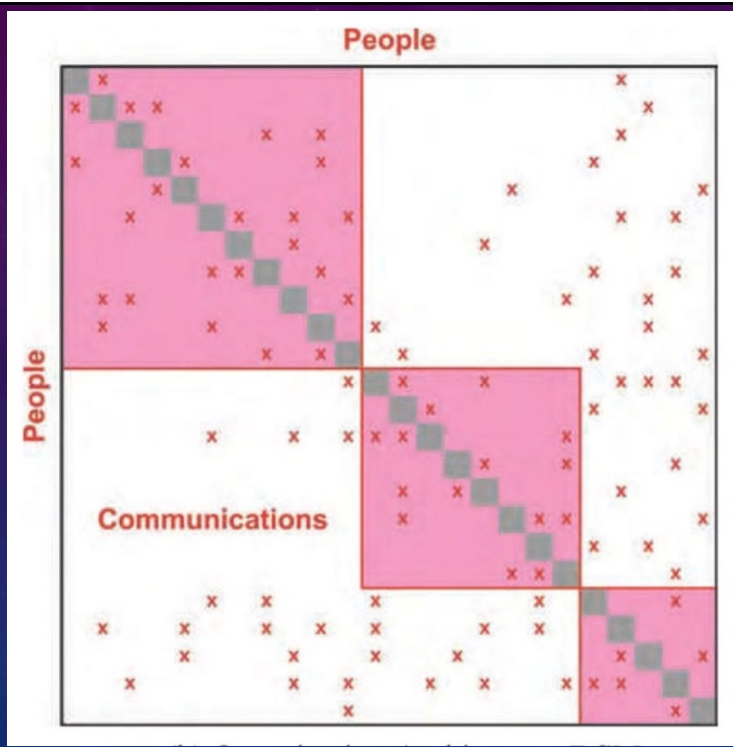
19

DSMs



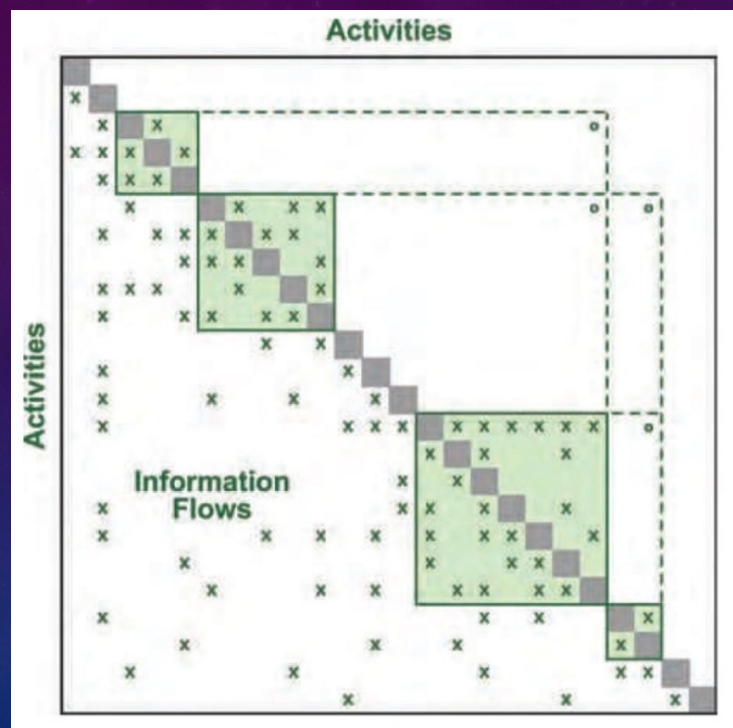
20

DSMs



21

DSMs



22

Step 2: Automated Architecture Analysis



Step 2.1 Measure and Monitor

Compare, Contrast, and Monitor



Step 2.2: Pinpoint Hotspots

Detection and Visualization



Step 2.3: Quantify Design Debt

Costs and benefits

23

Step 2.1: Measure and Monitor



- ✓ **Decoupling Level (DL):**
an options-based metric, measuring the system's ability to generate options
- ✓ **Propagation Cost (PC):**
a DSM-based metric, measuring how tightly coupled a system is

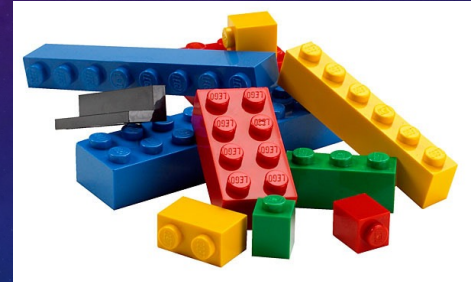
[Mo et al. ICSE 2016]

24

Decoupling Level (DL): Rationale

- 💡 A true module should be
 - Small
 - Independent

- 💡 A highly modularized system should
 - Have large numbers of true modules...
 - connected by design rules



[Mo et al. ICSE 2016]

25

Decoupling Level (DL)

The more files are clustered into true modules, the higher the value

Upper Layer modules:

- The fewer dependents, the higher the value
- The smaller the module, the higher the value

True modules:

- The smaller a true module, the higher the value
- The more true modules, the higher the value

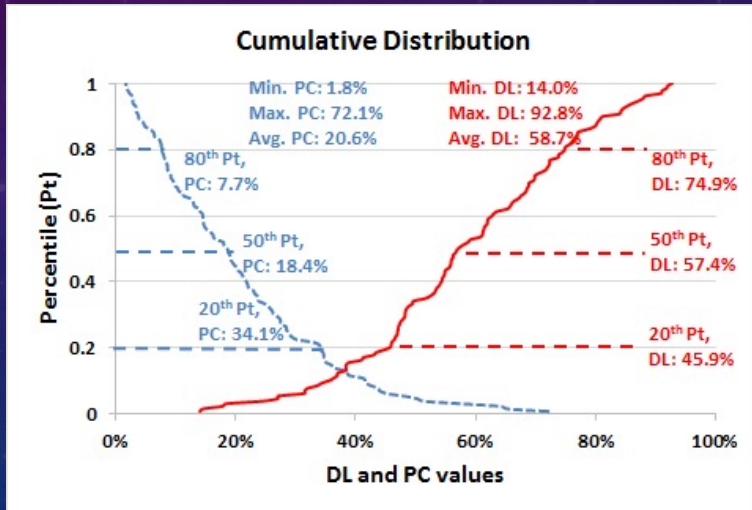
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1 UI_java	(1)																
2 Answer_java	x	(2)	x														
3 Question_java	x	x	(3)														
4 Survey_java				x	(4)												
5 SaveLoadFile_java						x	(5)										
6 TextFileUI_java	x						(6)										
7 CommandLineUI_java	x						(7)										
8 Letters_java								(8)									
9 Match_java	x	x						x	(9)	x							
10 MatchingAnswer_java	x	x	x					x	x	(10)							
11 EssayAnswer_java	x	x	x								(11)	x					
12 Choice_java	x	x	x								x	(12)					
13 EssayAnswer_java	x	x	x										(13)	x			
14 Written_java	x	x	x										x	(14)			
15 Test_java		x	x	x											(15)		
16 AnswerSheet_java																x	(16)

[Mo et al. ICSE 2016]

26

Decoupling Level (DL) and Propagation Cost (PC)

- Data from 129 projects:
- 108 open source
 - 21 industrial



[Mo et al. ICSE 2016]

27

DL and PC "Health Chart"

	Open Source		Commercial	
	DL	PC	DL	PC
Avg	60	20	54	21
Median	58	18	56	20
Max	92	72	93	50
Min	14	2	15	2
20 th Pt	47	8	36	6
40 th Pt	55	14	46	17
60 th Pt	66	21	59	24
80 th Pt	75	34	65	35

Pt: Percentile

Best DL (93%) is from industry
 Worst DL (14%) is from open source

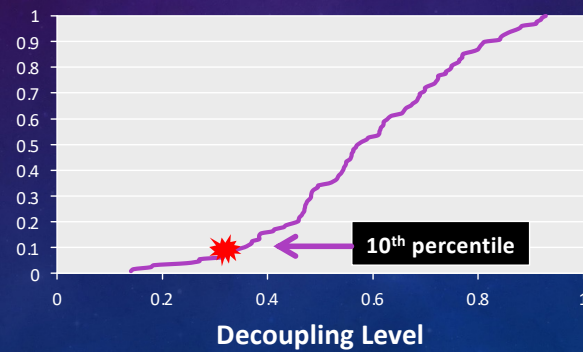
28

DL and PC "Health Chart"



An industrial project:

DL: 29%, 10th percentile: Confirmed to have severe maintenance difficulty



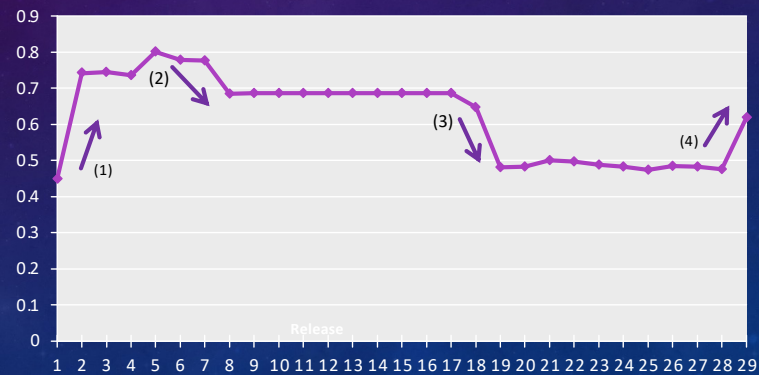
29

Evolutionary Monitoring



Non-trivial variation in DL indicates major architecture variation

- An industrial project: 6 years, 29 releases, 1082-1852 files



30

Step 2.2: Flaw Detection



- ✓ We automatically identify 6 types of design flaws
 1. Unstable interface
 2. Modularity violation
 3. Crossing
 4. Improper inheritance
 5. Cliques among files
 6. Package cycles
- ✓ These flaws are highly correlated with bugs, changes, and churn

[Mo et al. WICSA 2015]

31

Flaw Type 1: Unstable Interface

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 config.DatabaseDescriptor	(1)	dp,44	,14	,10	,10	,6	,14	,36	,118	,12	,	,16	,12	,42	,52	,4	,18	,30
2 utils.FBUilities	dp,44	(2)	,40	,4	,6	,10	,6	,12	,38	,28	,12	,8	,14	,24	,46	,6	,18	,28
3 utils.ByteBufferUtil	,14	dp,40	(3)	,	,	,	,	,4	,10	,20	,4	,4	,10	,26	,6	,12	,4	
4 service.WriteResponseHandler	,10	dp,4	,2	(4)	,4	,6	,18	dp,22	,	,	,	,	,	,6	,	,	,	
5 locator.TokenMetadata	,10	,6	,	,4	(5)	,4	,10	dp,24	,	,8	,	,	,	,4	,6	,4	,	
6 locator.NetworkTopologyStrategy	,6	dp,10	,2	,6	dp,4	(6)	,10	ih,22	,4	,	,	,	,	,16	,	,	,8	
7 service.DatacenterWriteResponseHandler	dp,14	dp,6	,2	ih,18	,10	dp,10	(7)	,20	,	,	,	,	,	,6	,6	,	,	
8 locator.AbstractReplicationStrategy	,36	dp,12	,4	dp,22	ag,24	,22	dp,20	(8)	,6	,	,	,	,	,16	,10	,	,10	
9 config.CFMetaData	,118	dp,38	dp,10	,	,4	,	,6	(9)	,	,	,	,16	,36	,46	,46	,	,56	
10 dht.RandomPartitioner	,12	dp,28	dp,20	,8	,	,	,	,	(10)	dp,4	,	,	,4	,16	,50	,	,	
11 utils.GuidGenerator	,	dp,12	,4	,	,	,	,	,	,4	(11)	,	,	,	,4	,	,	,	
12 io.sstable.SSTable	,16	,8	dp,4	,	,	,	,	ag,16	,	,	(12)	,4	dp,68	,10	,	,	,	
13 utils.CLibrary	,12	dp,14	,	,	,	,	,	,	,	,	,4	(13)	,12	,	,	,	,	
14 io.sstable.SSTableReader	dp,42	,24	dp,10	,	,	,	,	,36	,4	,	ih,68	dp,12	(14)	,22	,4	,	,10	
15 cli.CliClient	,52	dp,46	dp,26	,6	,4	,16	,6	,16	,46	,16	,4	,10	,	,22	(15)	,6	,14	,48
16 locator.PropertyFileSnitch	,4	dp,6	,	dp,6	,	,6	,10	,	,	,	,	,	,4	,6	(16)	,	,4	
17 dht.OrderPreservingPartitioner	dp,18	dp,18	dp,12	,4	,	,	,	,	,50	,	,	,	,	,14	,	(17)	,	
18 thrift.ThriftValidation	dp,30	,28	dp,4	,	,8	,	dp,10	dp,56	,	,	,	,10	,48	,4	,		(18)	

32

Flaw Type 2: Crossing

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	path1.File1_h	(1)	, 2		, 6	, 11	, 3	, 2	, 8	, 6	, 5		, 2	
2	path1.File2_h	d, 2	(2)			, 2								
3	path1.File3_h	d		(3)		, 2			, 2				, 2	
4	path2.File1_h	, 6			(4)	, 5	, 2	, 2	, 5	, 4			, 2	
5	path2.File2_h	d, 11	d, 2	d, 2	d, 5	(5)	, 3	, 2	, 10	, 6	, 5	, 2	, 3	, 2
6	path2.File3_h	d, 3			, 2	d, 3	(6)	d, 2	, 2	, 3	, 2		, 2	
7	path3.File1_h	d, 2			, 2	d, 2	d, 2	(7)	, 2	, 2	, 2		, 2	
8	path3.File2_c	d, 8		, 2	, 5	d, 10	d, 2	d, 2	(8)	, 6	d, 5		, 3	
9	path3.File3_c	d, 6			, 5	d, 6	d, 3	d, 2	d, 6	(9)	, 8		, 2	
10	path4.File1_c	d, 5			, 4	d, 5	d, 2	d, 2	d, 5	d, 8	(10)		, 2	
11	path4.File2_c	d				d, 2						(11)	, 7	
12	path4.File3_c	d, 2		, 2	, 2	d, 3	d, 2	d, 2	, 3	, 2	, 2			(12)
13	path5.File1_c	d	d			d, 2	d	d	d			d, 7	d	(13)

33

Flaw Type 3: Modularity Violation

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27			
1	path1.Bean.java	(1)																													
2	path2.Pear.java		(2)																												
3	path3.FirstFruit.java			(3)																											
4	path4.SecondFruit.java	Create,			(4)																										
5	path4.ThirdFruit.java	Create,				(5)																									
6	path4.FourthFruit.java						(6)																								
7	path5.FifthFruit.java							(7)																							
8	path5.Pear.java								(8)																						
9	path5.FirstJuice.java									(9)																					
10	path5.SecondJuice.java										(10)																				
11	path5.SixthFruit.java											(11)																			
12	path5.SeventhFruit.java												(12)																		
13	path5.EighthFruit.java														(13)																
14	path5.NinethFruit.java															(14)															
15	path5.TenthFruit.java																(15)														
16	path5.EleventhFruit.java																	(16)													
17	path5.TwelfthFruit.java																		(17)												
18	path5.ThirteenthFruit.java																			(18)											
19	path5.FourteenthFruit.java																				(19)										
20	path6.FirstFood.java																					(20)									
21	path7.SecondFood.java																						(21)								
22	path8.ThirdFood.java																							(22)							
23	path7.Apple.java																								(23)						
24	path9.FourthFood.java	Create,																								(24)					
25	path9.FifthFood.java	Cast,Use,																									(25)				
26	path10.FirstFig.java																											(26)			
27	path2.SecondFig.java	Cast,Use,																											(27)		

34

Sample Flaws from JBoss

1 \$RelationDataManager	(1)																			
2 JDBCEntityBridge	,35	(2)		ag,35																
3 CascadeDeleteStrategy			ag,	(3)	ag,															
4 JBDBCCMRFieldBridge	ag,163	ag,35	ag,	(4)																
5 JDBCInsertRelationsCommand	,15	,15		,15	(5)															
6 JDBCDeleteRelationsCommand		,11			,15	(6)														
7 JDBCPostCreateEntityCommand	,13	ag,		ag,13		(7)														
8 JDBCStopCommand	,11	ag,11		,11	,13	,11		(8)											ag,14	
9 JDBCRemoveEntityCommand	,21	ag,17		,21	,15	,13		,12	(9)										ag,20	
10 JDBCStartCommand	,18	ag,22		,18	,19	,15		,19	,16	(10)									ag,17	
11 JDBCLoadRelationCommand	,18	ag,21		,18	,19	,15	,11	,15	,15	,22	(11)								ag,19	
12 ReadAheadCache	,11			,11						,14	(12)								ag,	
13 JDBCStoreManager	,30	ag,23		,30	ag,18	ag,13	ag,	ag,14	ag,20	ag,17	ag,19	ag,	(13)							
14 JDBCAbstractQueryCommand	,17	ag,20		,17	,13	,11		,11	,13	,14	,23		(14)						ag,16	
15 JDBCCEJQLCompiler	,15	,14		,15	,12					,16	,22		ag,							26
16 RelationData				ag,																
17 RelationPair				ag,																
18 \$LeftJoinCMRNode	,17	,20		ag,17	,13	,11		,11	,13	,14	,23		,16	56	,26				(18)	,11
19 RelationSet	,16	,11		ag,16																(19)
20 SCMRChainLink	,32	,26		ag,32	,11		,12		,15	,14	,12		,16	,11					,11	(20)

RelationDataManager is in a modularity violation with the command classes

4 command classes and cache class also aggregate JDBCStoreManager

JDBCStoreManager aggregates command classes and cache class

Jboss JBDBCCMRFieldBridge DRSpace

35

Sample Flaws from Cassandra

1 cassandra.config.DatabaseDescriptor	(1)	dp,44	,14	,10	,10	,6	,14	,36	,118		,12	,12	,16	,42	,52	,30	,18	,4	
2 cassandra.util.FBUtilities	dp,44	2)	,40	,4	,6	,10	,6	,12	,38	,12	,28	,14	,8	,24	,46	,28	,18	,6	
3 cassandra.util.ByteBufferUtil	,14	dp,40	(3)					,4	,10	,4	,20		,4	,10	,26	,4	,12		
4 cassandra.service.WriteResponseHandler	,10	dp,4		(4)	,4	,6	,18	dp,22							,6				
5 cassandra.locator.TokenMetadata	,10	6		,4	(5)	,4	,10	dp,24			,8				,4		,4	,6	
6 cassandra.locator.NetworkTopologyStrategy	,6	dp,10		,6	dp,4	(6)	,10	ih,22	,4						,16	,8			
7 cassandra.service.DatacenterWriteResponseHandler	dp,14	dp,6		ih,18	,10	dp,10	(7)	,20							,6			,6	
8 cassandra.locator.AbstractReplicationStrategy	,36	dp,12	,4	dp,22	ag,24	,22	dp,20	(8)	,6						,16	,10	,10		
9 cassandra.config.CFMetaData	,118	dp,38	dp,10			,4		,6					,16	,36	,46	,56			
10 cassandra.util.GuidGenerator		dp,12	,4							(10)	,4				,4				
11 cassandra.dht.RandomPartitioner	,12	dp,28	dp,20		,8					dp,4	(11)			,4	,16		,50		
12 cassandra.util.CLibrary	,12	dp,14									(12)	,4	,12						
13 cassandra.io.sstable.SSTable	,16	8	dp,4						ag,16			,4	(13)	dp,68	,10				
14 cassandra.io.sstable.SSTableReader	dp,42	24	dp,10					,36		,4	dp,12	ih,68	(14)	,22	,10	,4			
15 cassandra.cli.CliClient	,52	dp,46	dp,26	,6	,4	,16	,6	,16	,46	,4	,16		,10	,22	(15)	,48	,4	,6	
16 cassandra.thrift.ThriftValidation	dp,30	28	dp,4			,8		dp,10	dp,56					,10	48	(16)		,4	
17 cassandra.dht.OrderPreservingPartitioner	dp,18	dp,18	dp,12		,4						,50				,14		(17)		
18 cassandra.locator.PropertyFileSnitch	,4	dp,6			dp,6		,6	,10						,4	,6	,4			(18)

36

Do Design Flaws Really Matter?

Research Question: If a file is involved in greater numbers of architecture issues, it is more error-prone/change-prone than average files?

[Mo et al. WICSA 2015]

37

Data Set

Subject	#Month	#Snap	#Comm	#Issue	#File
Avro 1.7.6	47	22	1480	630	145-298
Camel 2.11.1	53	46	17706	2326	528-1203
Cassandra 1.0.7	24	46	6738	3645	419-786
CXF 2.7.10	70	92	27247	3400	1426-3073
HBase 0.94.16	70	21	14858	5032	347-2142
Ivy 2.3.0	52	11	3799	839	418-607
OpenJPA 2.2.2	68	17	6736	1574	1216-1761
PDFBox 1.8.4	46	13	1798	1279	458-589
Wicket 1.5.5	57	55	18004	3359	1099-1549
Commercial	9	13	6000	800	137-599

38

Analysis

We counted the architecture flaws in these 10 projects and compared these to:

- Bug frequency
- Bug churn
- Change frequency
- Change churn

39

Results

Avro-1.7.6					Camel-2.11.1					Cassandra-1.0.7				
#AI	BF_avg	BC_avg	CF_avg	CC_avg	#AI	BF_avg	BC_avg	CF_avg	CC_avg	#AI	BF_avg	BC_avg	CF_avg	CC_avg
0	0.1	3.7	0.5	29.0	0	0.5	7.9	2.2	58.2	0	0.4	7.1	1.0	32.6
1	0.4	3.9	0.9	26.2	1	1.2	18.5	5.6	131.5	1	1.1	17.4	4.8	106.4
2	1.6	12.6	5.2	376.7	2	3.7	56.6	14.4	304.7	2	5.3	84.5	21.2	559.1
3	7.9	124.5	21.6	628.5	3	8.4	141.5	33.9	681.3	3	12.8	245.8	45.7	1202.0
4	16.5	255.0	33.5	1220.0	4	13.9	204.7	50.9	1043.5	4	18.8	364.9	65.7	1909.4
PC	0.91	0.89	0.94	0.95	PC	0.96	0.96	0.97	0.97	PC	0.97	0.96	0.98	0.97
CXF-2.7.10					Hadoop-2.2.0					HBase-0.94.16				
#AI	BF_avg	BC_avg	CF_avg	CC_avg	#AI	BF_avg	BC_avg	CF_avg	CC_avg	#AI	BF_avg	BC_avg	CF_avg	CC_avg
0	0.8	21.0	2.8	86.9	0	0.4	12.7	1.0	56.8	0	0.7	10.4	0.9	53.0
1	2.9	62.3	9.4	262.5	1	1.5	24.8	4.2	167.7	1	4.8	236.7	8.3	614.6
2	8.6	164.8	23.1	592.0	2	5.3	173.6	13.8	558.3	2	9.9	418.5	17.2	2083.6
3	20.2	390.9	52.5	1232.4	3	26.0	725.1	58.0	1959.6	3	47.8	1335.1	87.6	3158.7
4	54.1	890.2	142.3	3326.0	4	13.7	237.9	26.8	1252.0	4	76.7	2370.4	135.1	6019.0
PC	0.90	0.92	0.89	0.89	PC	0.76	0.63	0.72	0.83	PC	0.93	0.94	0.93	0.97
Ivy-2.3.0					OpenJPA-2.2.2					Pdfbox-1.8.4				
#AI	BF_avg	BC_avg	CF_avg	CC_avg	#AI	BF_avg	BC_avg	CF_avg	CC_avg	#AI	BF_avg	BC_avg	CF_avg	CC_avg
0	0.2	4.5	1.1	31.8	0	1.8	10.0	1.1	36.8	0	0.5	27.1	1.1	92.0
1	1.1	22.8	3.3	79.6	1	3.2	31.1	3.7	111.5	1	1.4	35.9	2.9	136.5
2	2.9	54.6	8.4	251.9	2	4.6	64.5	7.5	229.8	2	1.5	64.1	3.4	259.9
3	7.0	119.9	20.9	646.2	3	10.8	408.6	22.4	862.5	3	8.1	495.0	13.7	861.3
4	6.4	204.6	18.6	792.3	4	25.1	981.0	52.5	2301.1	4	12.2	669.5	18.4	1254.4
PC	0.94	0.96	0.93	0.97	PC	0.90	0.88	0.90	0.88	PC	0.92	0.92	0.94	0.94
Commercial Project														
#AI	BF_avg	BC_avg	CF_avg	CC_avg										
0	0.1	2.25	2.7	102.5										
1	0.2	4.6	5.9	200.4										
2	0.8	3.24	10.3	372.0										
3	2.8	36.8	19.8	884.7										
4	6.0	21	29.0	549.0										
PC	0.91	0.73	0.98	0.81										

40

Avro-1.7.6

#Flaws	BF_avg	BC_avg	CF_avg	CC_avg
0	0.1	3.7	0.5	29.0
1	0.4	3.9	0.9	26.2
2	1.6	12.6	5.2	376.7
3	7.9	124.5	21.6	628.5
4	16.5	255.0	33.5	1220.0
PC	0.91	0.89	0.94	0.95

41

More Consequences of Design Flaws

Research Question: If a file is involved in greater numbers of architecture flaws, it is involved in more *security* bugs/changes than average files?

[Feng et al. WICSA 2016]

42

Answer

We counted the architecture flaws in these 11 projects and compared these to:

- Security bug frequency
- Security change frequency
- ...as well as the original measures (bugs, changes, bug churn, change churn)

43

Answer

Project	Bug/Flaw Correlation	Change/Flaw Correlation	Security Bug/Flaw Correlation
Avro	0.845	0.923	0.861
Camel	0.956	0.959	0.958
Cassandra	0.830	0.869	0.808
Chrome	0.987	0.988	0.979
CXF	0.896	0.910	0.939
Derby	0.938	0.917	0.897
Hadoop	0.752	0.902	0.862
HBase	0.894	0.932	0.961
httpd	0.710	0.688	0.885
PHP	0.929	0.987	0.923
Tomcat	0.901	0.776	0.920

44

Step 2.3: Quantification



- ✓ Calculate the costs of each root, each flaw and each type of flaw
- ✓ Calculate ROI (Return on Investment)

[Kazman et al. ICSE 2015]

[Xiao et al. ICSE 2016]

45

Industrial Experience: ROI Calculation

Penalty Caused by Architecture Debt Refactoring Cost Expected Savings

	A	B	C	D	E	F	G	I	J	K	L	M	N
	DRSpace Leading File	DRSpace Size	Norm Size	Current Defects/Yr	Norm Defects	Current Changes/Yr	Norm Changes/Yr	Tot LOC Changed	Norm LOC Changed	Refactor Cost (PM)	Norm Exp Defects/Yr	Norm Exp Changes/Yr	Norm Exp LOC Changed
2	Pear.java	139	119.33	166	142.5	1068	839.2	49,171	42,213	5.5	39	346	20,281
3	Apple.java	158	133.83	63	53.4	607	451.7	25,603	21,686	7	44	388	22,745
4	Bean.java	65	37.83	72	41.9	429	207.2	17,807	10,364	1.5	12	110	6,429
6	DRSpace Total		290.99		237.8		1498		74,263		96.0	843.871	49,455
7	Project Total	797		265		2332		135,453		14			
8	Savings										142	654	24,808
11	Base defect rates	0.33											
12	Base change rates	2.9										Exp PM save	41.35
13	Base LOC/file	169.95											
14	LOC/PM	600											

Result: ~300% ROI in the first year alone!

[Kazman et al. ICSE 2015]

46

Industrial Experience: Analyzing 8 ABB Projects

- ✓ Using 3 complementary techniques:
 - Architecture-level maintainability metrics
 - Architecture flaw analysis
 - Cost and benefit analysis
- ✓ 8 projects developed at multiple locations (India, USA, Switzerland) differing in age, domain, and size.
- ✓ We reported the results back to each project and collected feedback

[Mo et al. ASE 2018]

47

Collecting Feedback from ABB

- ? RQ1: does the tool suite help to close the gap between management and development? That is, does it help them to decide *if*, *when*, and *where* to refactor?
- ? RQ2: does the tool suite help practitioners understand the maintainability of their systems *relative to other projects* internal to the company, and relative to a more broad-based benchmark suite?
- ? RQ3: does the tool suite help developers pinpoint the *hotspots* of their systems, i.e., the groups of files with severe design flaws?

48

ABB: Metrics Scores and Rankings

	DL	Percentile	PC	Percentile	#Files
Proj_EO	78%	85th	6%	85th	144
Proj_BM	77%	85th	2%	98th	371
Proj_CH	76%	81st	16%	54th	6,948
Proj_EP	72%	74th	7%	83th	1,541
Proj_SS	57%	49th	20%	45th	15,333
Proj_OP	57%	49th	21%	41th	7,754
Proj_CO	55%	43rd	17%	52th	491
Proj_EC	28%	5th	62%	2nd	4,125

49

ABB: Hotspots Detected

Unstable Interface

File Cliques

Modularity Violations

Most error-prone files

	CF	Top	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
1 p1.F1	17	3	(1)	d,6		4	2					6	4	2	8		2	3					3	2	2	3	9	2	2	2	2			
2 p1.F2	8	5	i,6	(2)	3	3	2					3		2	3								2	2	2	2	4							
3 p1.F3	16	3	i,6	(3)	(3)	2	2					3		2	3	3	3						2	2	2	3	4	3	2		3			
4 p1.F4	2	24				(4)								2												2								
5 p1.F5	45	1	d,4			(5)	4	4	d,4			5	7	2	3	2			2	2	3	4	4	12	2	6	5	3	4					
6 p1.F6	22	2	d,2			d,4	(6)	2					10	2								3	3	3	6	d,7	3	2						
7 p1.F7	1	43	i,			(7)																				2	2	8	4		2			
8 p2.F1	18	3	d			4	d,2	d	(8)	5		5													2	2	8	4		2				
9 p2.F2	6	6	d			d,4		d,5	(9)																	2	2	2						
10 p2.F3	10	4	i,6	(3)	3	3	2				(10)	3		2	5		2					2	2	2	2	6			2			2		
11 p2.F4	5	8	d,4								i,3	(11)			4											5								
12 p2.F5	2	24										(12)														2								
13 p2.F6	30	2	d,2		2	i,2	d,5	d,10	d,5	d,5	2		d	(13)	2				2				3	4	4	8	11	3						
14 p2.F7	1	43																																
15 p2.F8	40	1	d,8			d,7	2				5	4	2	d	(14)	4	4					2	3	3	4	30	6	3	3	4				
16 p3.F1	11	4	d			3	2									4	(16)	3								5	3	2		3				
17 p3.F2	2	24	i,2																															
18 p3.F3	15	3	d,3			3	3	2			2		2	4	3		(18)									3	3	2	2	3				
19 p3.F4	41	1	d,																			(19)	20	8	2	2	26	38	3					
20 p3.F5	4	9																					(20)			4								
21 p3.F6	82	0,3	d,																			d,2	d	(21)	2	2	16	59	3					
22 p3.F7	28	2	d,			2																d,8	2	(22)	6	2	6	25	5		2			
23 p3.F8	10	4	d,3			2	3	3			2		3	2	2							d,2	i,6	(23)	2	5	10	3						
24 p4.F1	14	3	d,2			2	d,4	3			2		4	3								2	2	d,2	2	2	(24)	7	8	4				
25 p4.F2	59	1	d,2			2	d	d,4	3	2		2	4	3								i,26	d,16	5	7	(25)	42	5						
26 p4.F3	361	0,1	d,3			3	d,2	d,12	6	2		2	8	4	3							d,3	d,4	d,59	i,25	10	d,8	d,42	(26)	21	2	2	3	
27 p4.F4	114	0,2	d,9	4	d,4	d,2	d,7	d	d,8	d,2	d,6	d,5	2	d,11	d,3	d,5	d	d,6	d,3	2		d	3	3	4	3	d,21	(27)	14	8	6	5		
28 p4.F5	29	2	d,2			d,3	d,6	3	d,4	d,2	d	d	3									d					2	d,14	(28)	14	2	7		
29 p4.F6	21	2	d,2			2	d,5	2																			2	8	d,1	(29)	2	7		
30 p4.F7	10	4	d,2			3																				3	d,6	2	2	(30)	2			
31 p4.F8	12	4	d,2			3	4		2	2								4	3	3							5	d,7	7	2	(31)			

50

ABB: Collecting User feedback



✓ Surveys

✓ Interviews

51

ABB Post-Mortem: Surveys with Architects

Q1: What did the report reveal that you didn't know about your software?

Q2: Are the metrics useful for reflecting the architecture of your software?

Q3: What did the architecture design flaws reveal about your software?

Q4: What actions have you planned as a result of the architecture design flaws report?

Q5: What did the architecture roots reveal about your software?

Q6: What actions do you plan to take to address architecture roots?

52

Results

RQ1: does the tool suite help to close the gap between management and development? That is, does it help them to decide if, when, and where to refactor?



Participants of all 8 projects verified that the information provided was useful in closing the understanding gap with management. They have begun the refactoring process.

53

Results

RQ2: does the tool suite help practitioners understand the maintainability of their systems relative to other projects internal to the company, and relative to a more broad- based benchmark suite?



All participants said the report gave them quantifiable results with which to judge their project. The comparison with industrial benchmarks made it clear that maintenance difficulty caused by degrading architecture is common.

54

Results

RQ3: does the tool suite help developers pinpoint the *hotspots* of their systems—that is, the groups of files with severe design flaws?

- ✓ Six of the eight projects planned to or already started refactoring to address the detected flaws. The project with the lowest DL score is undergoing a major rewrite.

55

Industrial Experience: Huawei

- Developed a set of architecture measures based on DL and architecture flaws
 - Adopted as a corporate standard
 - Now used in over 100 projects
- Quantified architecture debt
- 24 out of 29 projects studied showed a positive correlation between these measures and productivity

[Wu et al. ECSA 2018]

56

Industrial Experience: BrightSquid

- Analyzed BrightSquid's secure communication platform (6/16 – 5/17)
- Identified many areas of architecture debt—the "before" state—and recommended a refactoring plan to pay down the debt (7/17)
- Architecture was refactored (1/18 – 3/18)
- Analyzed the "after" state (3/18 – 8/18)

[Nayebi et al. ICSE 2019]

57

BrightSquid Results

General information	Before	After
# of files	1713	711
# of roots covering 80% of bugs	5	3
# of files in roots covering 80% of bugs	296	295
# of files covering 80% of bugs	17%	37%
Architectural Metrics	Before	After
Decoupling level	86%	83%
Propagation cost	6%	6%
Architectural flaws	Before	After
# of cliques	17	10
# of files influenced by cliques	71	26
# of unhealthy inheritance	60	30
# of files influenced by unhealthy inheritance	222	102
# of unstable interface	12	8
# of files influenced by unstable interface	471	59
# of crossings	29	6
# of files influenced by crossings	387	47
# of package cycles	34	19
# of files influenced by package cycles	242	94

58

Industrial Experience: BrightSquid

- The refactoring activities were recorded as 106 change requests, which consumed 563.8 person hours.
- After refactoring, the size of the code base shrunk by 41.5%
- The average time needed to close issues before and after refactoring was reduced by 72%.
- The average bug-fixing churn per issue dropped by 2/3: from 102 LOC before refactoring to 34 LOC after refactoring
- The average bug-fixing duration reduced 30%, dropping from 10 days before to 7 days

59

Lessons Learned

- There is enormous design debt in today's software.
- Yes, in your software.
- That's the bad news.
- The good news: we can do something about it.

60

Lessons Learned

- The good news: It is possible to automatically and objectively assess and quantify architecture quality – to find and fix the debt.
- And it is possible to bridge the gap. These results were enthusiastically received by the industrial projects.
- Most projects are embarking on major refactorings.
- Several companies have incorporated DV8 into their development processes/pipelines.

61

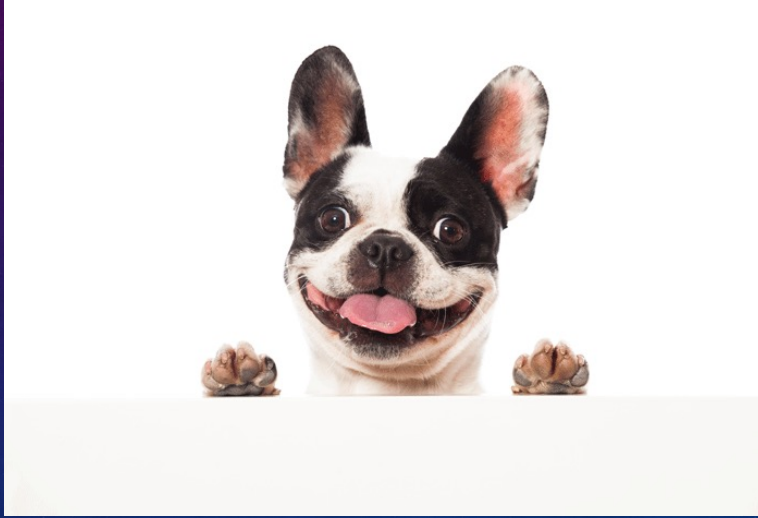
Final Thoughts

- ✓ You can't manage it if you don't measure it. Quantification is key.
- ✓ If the measurement is not automated it won't be done, or won't be repeatable.
- ✓ Incorporating these techniques into the build process ensures rapid feedback with supporting data.
- ✓ This measurement, detection, and quantification practice leads to improved architectures.
- ✓ Results must be accompanied by ROI measures, to aid in adoption.

You can get the software—free for academic use—at: <https://archdia.com/>

62

Thank You!



63



64

Acknowledgments

- I owe a huge debt of gratitude to my co-authors, without whom none of this would be possible.
- Special thanks to:
 - Yuanfang Cai
 - Lu Xiao
 - Ran Mo
 - Qiong Feng
 - Jason Lefever
 - Humberto Cervantes



65

References

[Lefever et al. 2021] Jason Lefever, Yuanfang Cai, Humberto Cervantes, Rick Kazman, Hongzhou Fang, "On the Lack of Consensus Among Technical Debt Detection Tools", ICSE 2021 SEIP

[Xiao et al. 2014] Lu Xiao, Yuanfang Cai, Rick Kazman, "Design rule spaces: a new form of architecture insight", ICSE 2014: 967-977

[Kazman et al. 2015] Rick Kazman, Yuanfang Cai, Ran Mo, Qiong Feng, Lu Xiao, Serge Haziyeve, Volodymyr Fedak, Andriy Shapochka, "A Case Study in Locating the Architectural Roots of Technical Debt", ICSE 2015: 179-188

[Mo et al. 2016] Ran Mo, Yuanfang Cai, Rick Kazman, Lu Xiao, Qiong Feng, "Decoupling level: a new metric for architectural maintenance complexity", ICSE 2016: 499-510

[Mo et al. 2018] Ran Mo, Will Snipes, Yuanfang Cai, Srinivas Ramaswamy, Rick Kazman, Martin Naedele: "Experiences Applying Automated Architecture Analysis Tool Suites", ASE 2018.

[Mo et al. 2019] Ran Mo, Yuanfang Cai, Rick Kazman, Lu Xiao, and Qiong Feng: "Architecture Anti-patterns: Automatically Detectable Violations of Design Principles", IEEE Transactions on Software Engineering, 2019

[Nayebi et al. 2019] Maleknaz Nayebi, Yuanfang Cai, Rick Kazman, Guenther Ruhe, Qiong Feng, Chris Carlson, Francis Chew: "A Longitudinal Study of Identifying and Paying Down Architectural Debt", ICSE SEIP 2019.

[Feng et al. 2019] Q. Feng, Y. Cai, R. Kazman, D. Cui, T. Liu, H. Fang. "Active Hotspot: An Issue-Oriented Model to Monitor Software Evolution and Degradation", ASE 2019.

66

66