



Intelligent System  
Logging and Monitoring  
ISyLM Research Lab

# Using AI to Empower System Development and Operations Team

**Wahab Hamou-Lhadj, PhD**  
ECE, Concordia University  
[wahab.hamou-lhadj@concordia.ca](mailto:wahab.hamou-lhadj@concordia.ca)

TNSBC, Montréal, QC  
May 23, 2018

# Software-intensive systems are everywhere!

- Health
- Energy
- Finance
- Manufacturing
- Education
- Public safety
- Telecom
- Aerospace
- Entertainment
- Hospitality
- Public administration
- Social interactions

***“Our civilization  
runs on software”***

*B. Stroustrup*

# Facts

- From 1997 to 2012, software industry production grew from \$149 billion to \$425 billion.
- The software industry's direct share of U.S. GDP went from 1.7% to 2.6%.
- Software accounted for 12.1% percent of all U.S. labor productivity gains from 1995 to 2004 and 15.4% from 2004 to 2012.

## The U.S. Software Industry: An Engine for Economic Growth and Employment

---

Software & Information  
Industry Association  
[www.siaa.net](http://www.siaa.net)



DEVELOPED FOR THE PUBLIC POLICY DIVISION OF THE  
SOFTWARE & INFORMATION INDUSTRY ASSOCIATION (SIIA)

By Robert J. Shapiro of Sonecon

# Software Development

A set of activities for creating a software system including requirements analysis, architectural design, detailed design, coding, testing, maintenance, integration, acceptance testing, etc.

# Software Operations

A set of activities for supporting end users of a software product in an operational environment. Typical activities include: installation, upgrade, monitoring, configuration, etc.

Source: Definitions adapted from ISO/IEC24748-1 2011, ISO/IEC15288 2008, ISO/IEC12207 2008).

# Software Development

A set of activities for creating a software system including requirements analysis, architectural design, detailed design, coding, testing, maintenance, integration, acceptance testing, etc.

# Software Operations

A set of activities for supporting end users of a software product in an operational environment. Typical activities include: installation, upgrade, monitoring, configuration, etc.

Source: Definitions adapted from ISO/IEC24748-1 2011, ISO/IEC15288 2008, ISO/IEC12207 2008).

# SW Development Challenges

- Increased complexity
- High cost
- Heavy reliance on people
- Lack of automated tools
- Time to market pressure
- Maintaining quality



A NIST study\* shows that defects in software cost the U.S. economy **\$56 billion annually**.

A large percentage of software development costs are **spent on identifying and correcting defects**.

There is a need to **invest in automated and intelligent solutions**.

\*Source: Research Triangle Institute, *The Economic Impacts of Inadequate Infrastructure for Software Testing*, NIST Planning Report 02-3, May 2002.

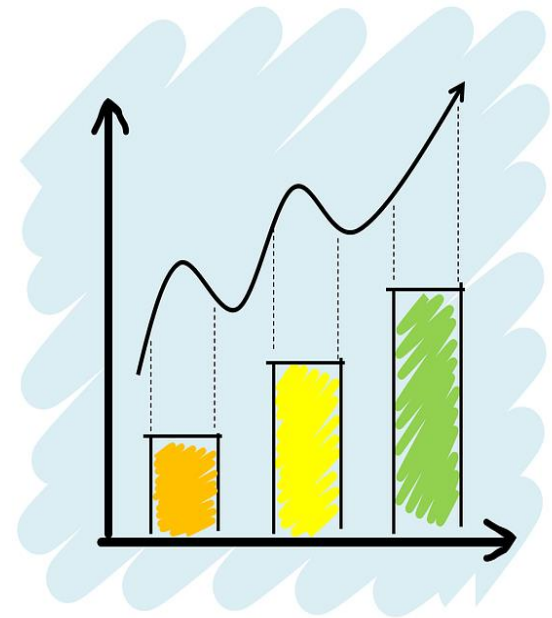
# Active Research Community

- Change and defect management
- Continuous integration/deployment
- Release engineering
- Reverse engineering and re-engineering
- Run-time evolution and dynamic configuration
- Software and system comprehension
- Software migration and renovation
- Software refactoring and restructuring
- Empirical studies
- Evolution of non-code artefacts
- Human aspects of software maintenance and evolution
- Maintenance and evolution of model-based methods
- Maintenance and evolution processes
- Maintenance and evolution of mobile apps
- Maintenance versus release process
- Mining software repositories
- Etc.



# Emergence of Software Analytics

- Data-driven software development and maintenance
- Big Data: source code, bug reports, test cases, logs, user feedback, etc.
- Predictive analytics using ML, DL, CI, and PR
- Information visualization of large-scale data

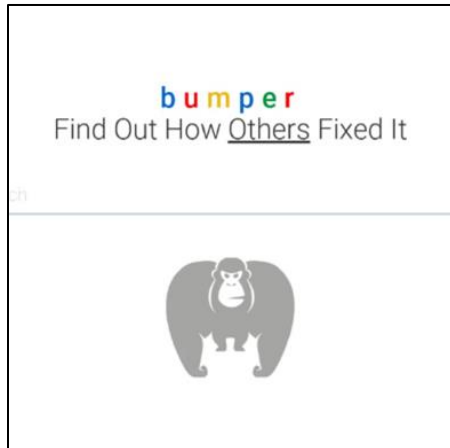


# The Commit Assistant Project

- An NSERC project in collaboration with Ubisoft.
- Goal: To empower SW developers with an intelligent tool that detects defects as they write code, and proposes fixes.



# Open Technologies behind CommitAssistant



Bug Metarepository  
Search Engine for  
Developers and  
Reseachers

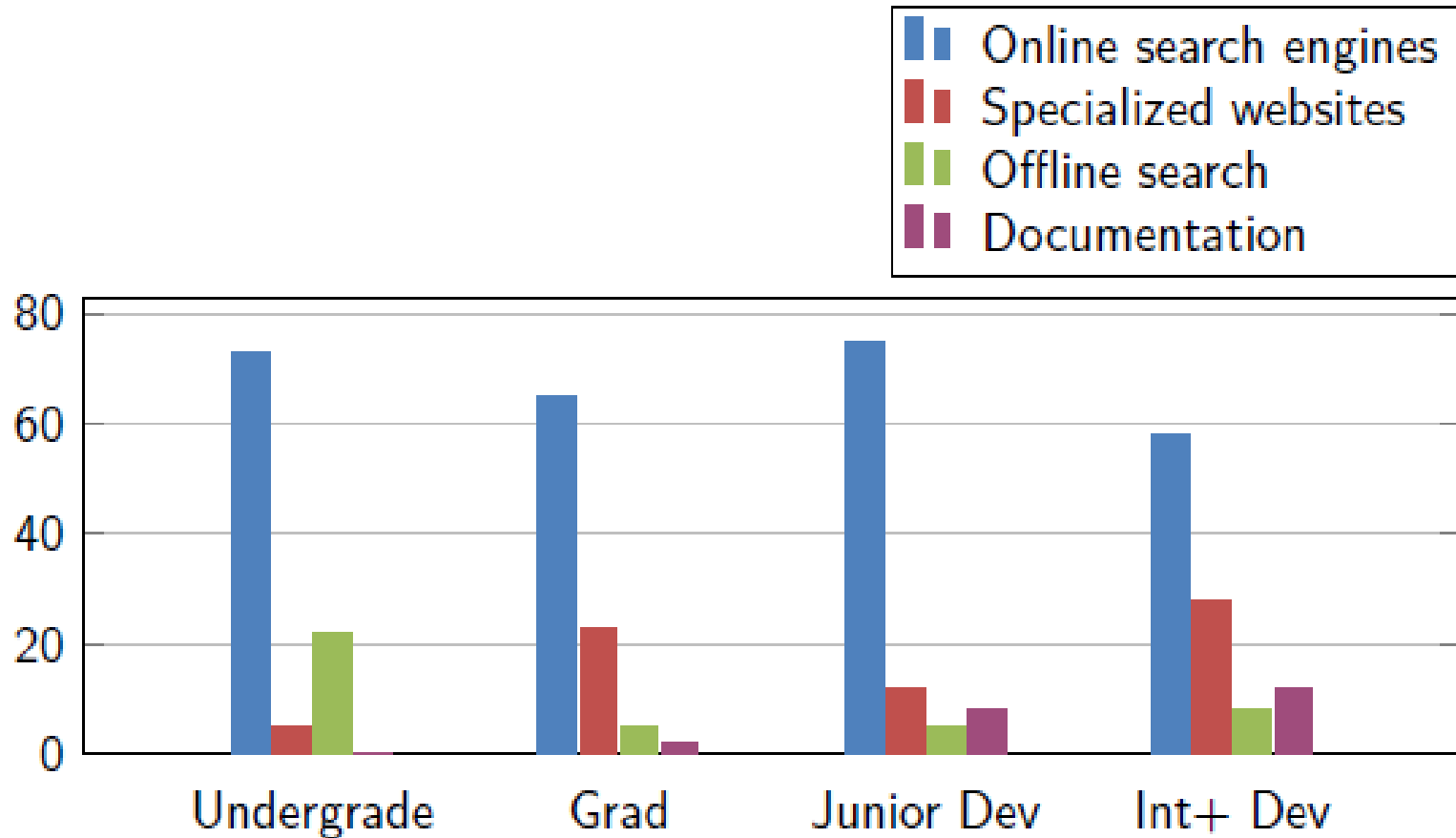
**BIANCA**

Preventing Bug  
Insertion at Commit-  
Time Using Clone  
Detection

**CLEVER**

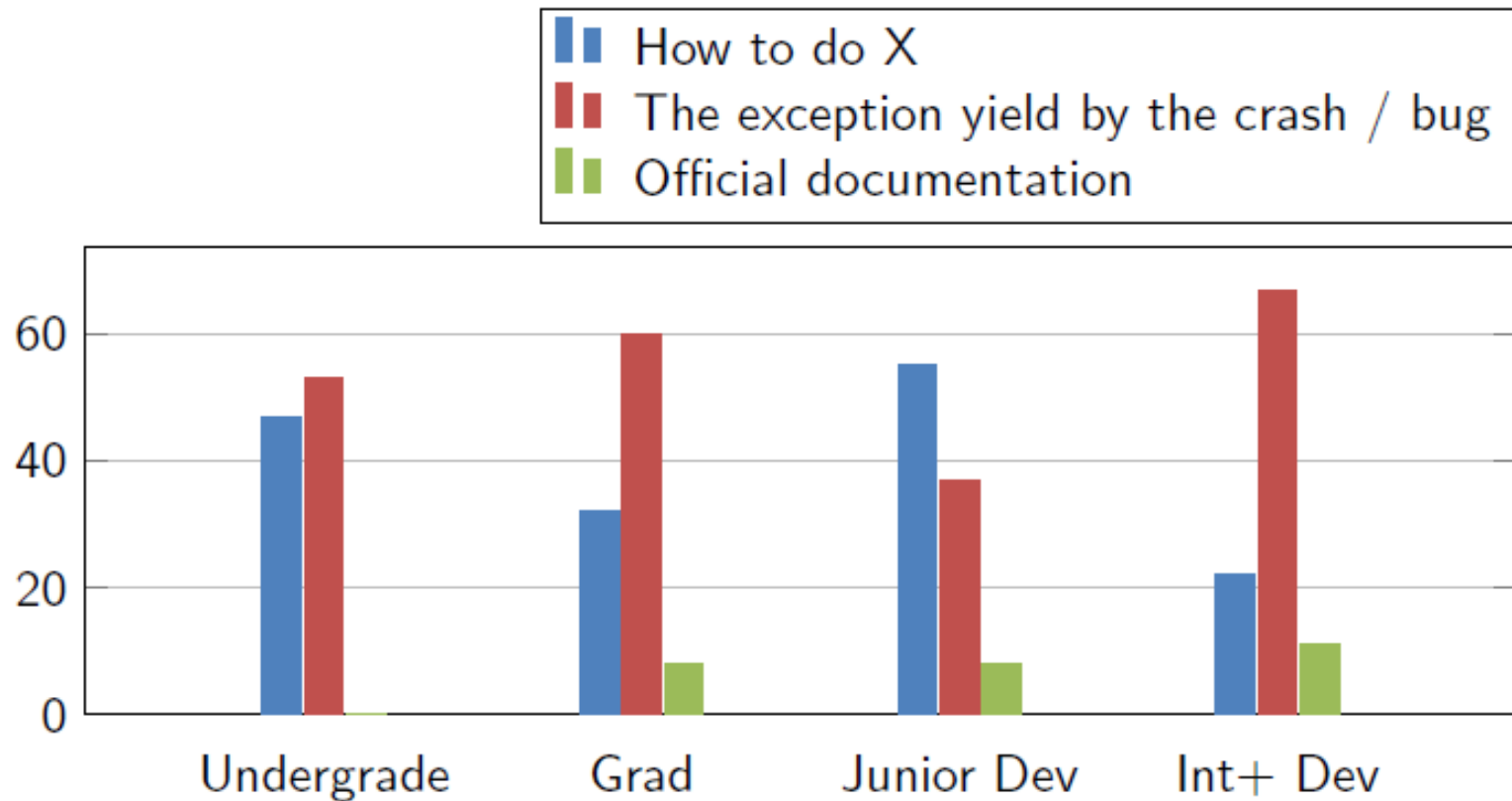
Combining Levels of  
Bug Prevention and  
Resolution  
Techniques

# Where do developers look for information when facing an unknown bug/crash?



89 participants

# What do developers search for when facing an unknown bug/crash?



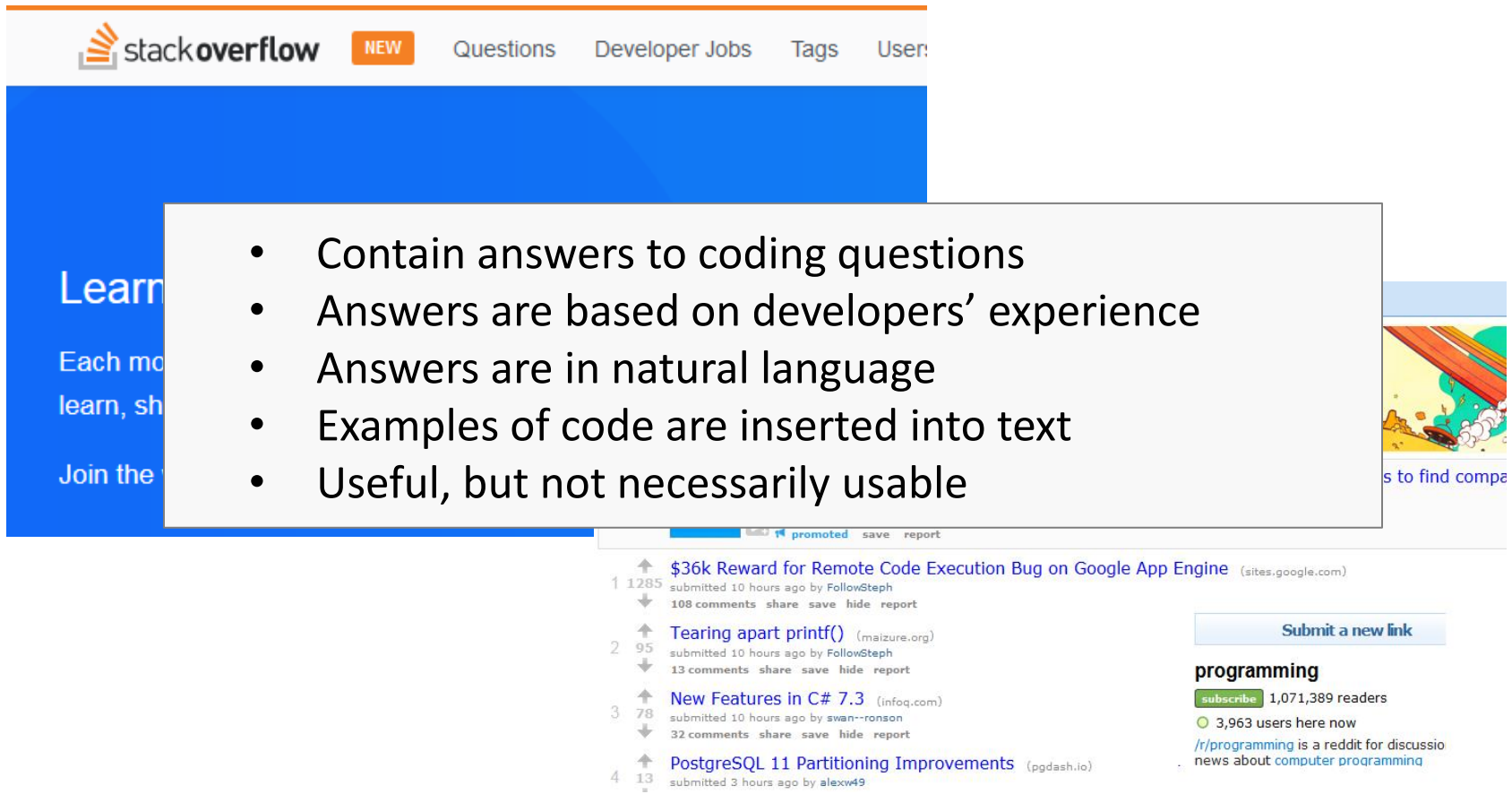
89 participants

# Examples of Coding Websites

The image shows the top portion of the Stack Overflow website. At the top left is the Stack Overflow logo, followed by a 'NEW' button and navigation links for 'Questions', 'Developer Jobs', 'Tags', and 'Users'. Below this is a large blue banner with the text 'Learn, Share, Build' in white. A red circle highlights the word 'Build'. Below the banner, there is more text: 'Each month, over 50 million developers come to learn, share their knowledge, and build their skills.' and 'Join the world's largest developer community.'

The image shows the top portion of the Reddit website. At the top left is the Reddit logo, followed by the word 'PROGRAMMING' and a 'hot' button. Other navigation links include 'new', 'rising', 'controversial', 'top', 'gilded', and 'wiki'. Below this is a yellow banner with the text 'Welcome to Reddit. Come for the cats, stay for the empathy.' and a 'BECOME A REDDITOR' button. Below the banner is a list of posts. The first post is a promoted link: 'The hiring process is a Bayesian classification problem. Use our trained models to find companies you're most likely to get offers.' (triplebyte.com). The second post is '\$36k Reward for Remote Code Execution Bug on Google App Engine' (sites.google.com) with 1285 upvotes and 108 comments. The third post is 'Tearing apart printf()' (maizure.org) with 95 upvotes and 13 comments. The fourth post is 'New Features in C# 7.3' (infoq.com) with 78 upvotes and 32 comments. The fifth post is 'PostgreSQL 11 Partitioning Improvements' (pgdash.io) with 13 upvotes. On the right side, there is a 'Submit a new link' button and a 'programming' subreddit card with 1,071,389 readers and 3,963 users here now. A red circle highlights the 'Submit a new link' button and the 'programming' subreddit card.

# Examples of Coding Websites



The screenshot shows the Stack Overflow website interface. At the top, there is a navigation bar with the Stack Overflow logo, a 'NEW' badge, and links for 'Questions', 'Developer Jobs', 'Tags', and 'Users'. Below the navigation bar is a blue sidebar on the left with the text 'Learn', 'Each mo', 'learn, sh', and 'Join the'. The main content area displays a list of questions with their respective scores, submission times, and authors. A 'Submit a new link' button is visible on the right side of the page.

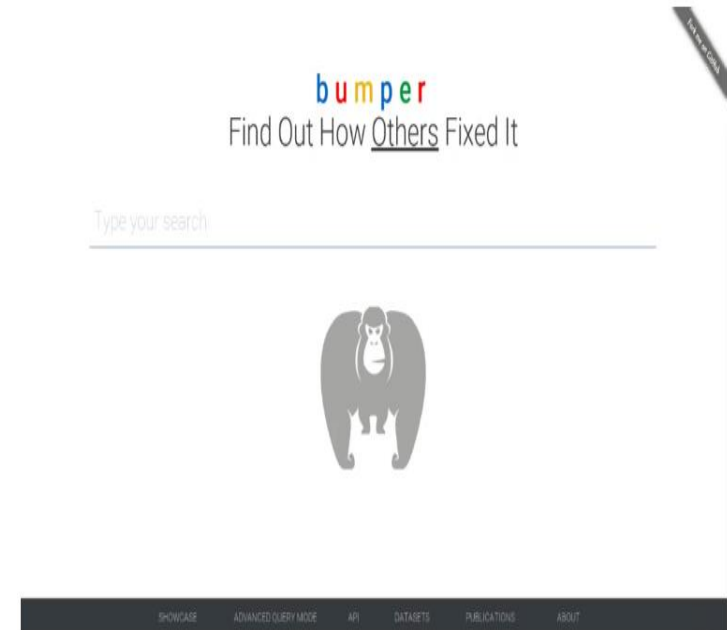
- Contain answers to coding questions
- Answers are based on developers' experience
- Answers are in natural language
- Examples of code are inserted into text
- Useful, but not necessarily usable

Rank	Score	Title	Author	Submitted
1	1285	<a href="#">\$36k Reward for Remote Code Execution Bug on Google App Engine</a>	(sites.google.com)	submitted 10 hours ago by FollowSteph
2	95	<a href="#">Tearing apart printf()</a>	(maizure.org)	submitted 10 hours ago by FollowSteph
3	78	<a href="#">New Features in C# 7.3</a>	(infoq.com)	submitted 10 hours ago by swan--ronson
4	13	<a href="#">PostgreSQL 11 Partitioning Improvements</a>	(pgdash.io)	submitted 3 hours ago by alexw49

**programming**  
subscribe 1,071,389 readers  
3,963 users here now  
[/r/programming](#) is a reddit for discussio  
news about computer programming

# BUMPER: Bug Metarepository Search Engine for Developers and Researchers

- Aggregates information from many bug report and code versioning systems
- Is an online search engine to millions of bug reports and fixes from open-source repositories
- Uses a query system for developers and advanced API for researchers
- Leverages the concept of collective coding → collective intelligence





# bumper

Find Out How Others Fixed It

Type your search

---



SHOWCASE

ADVANCED QUERY MODE

API

DATASETS

PUBLICATIONS

ABOUT

**Developers can search millions of lines of code and bug reports for a bug or crash they encountered.**

User query

# NullPointerException

About 27626 results (0.01 seconds)

LANGUAGES DATASETS

DOWNLOAD

▲ NullPointerException at org.netbeans.api.java.source.JavaSource\$JavaSourceAccessorImpl.  
 38 [https://netbeans.org/bugzilla/show\\_bug.cgi?id=189412](https://netbeans.org/bugzilla/show_bug.cgi?id=189412) java, netbeans, java  
 Build NetBeans IDE 6.9 (Build 201006101454) VM: Java HotSpot(TM) Client VM, 16.2-504, Java(TM) SE Runtime Environment, 1.6.0\_19-b04 OS: Windows 7 Backtrace: java.lang.NullPointerException

▲ NullPointerException at org.netbeans.modules.java.source.usages.LuceneIndex\$DirCache.cle  
 10 [https://netbeans.org/bugzilla/show\\_bug.cgi?id=189499](https://netbeans.org/bugzilla/show_bug.cgi?id=189499) java, netbeans, java  
 IDE Dev (Build 201008130001) VM: Oracle JRockit(R), R28.0.1-020100512-2131-windows-b6\_64, Java(TM) SE Runtime Environment, Windows

cting thread cpu timestamps by default  
[https://netbeans.org/bugzilla/show\\_bug.cgi?id=189821](https://netbeans.org/bugzilla/show_bug.cgi?id=189821) java, netbeans, profiler  
 that thread cpu timestamps are available to JVM [1.6+] on all platforms and obtaining them is reasonably quick [1] we can enable collecting

▲ NullPointerException at java.util.Arrays\$ArrayList.<init>  
 2 [https://netbeans.org/bugzilla/show\\_bug.cgi?id=177814](https://netbeans.org/bugzilla/show_bug.cgi?id=177814) java, netbeans, platform  
 Build NetBeans IDE Dev (Build 200911081400) VM: Java HotSpot(TM) Client VM, 6.0.5

176129.806c07f52485#189412 NullPointerException at org.netbeans.ap.java.source.JavaSource\$JavaSourceAccessorImpl.setJavaSource  
 • java source/src/org/netbeans/modules/java/source/JavaSourceAccessor.java  
 • java source/src/org/netbeans/modules/java/source/paring/JavaParserResult.java  
 (2) files, (10) insertions, (4) deletions.

```
Index: java_source/src/org/netbeans/modules/java/source/JavaSource
Accessor.java
-----
+
@@ -24,6 +24,7 @@
import org.netbeans.text.PositionRef;
import org.netbeans.util.Exceptions;
import org.netbeans.util.Mutex;
+import org.netbeans.util.Parameters;
/**
 *

```

```
Index: java_source/src/org/netbeans/modules/java/source
Accessor.java
-----
+
@@ -316,9 +317,14 @@
}
@Override
public void run(Result result, SchedulerEvent
public void run(@NotNull Result result, Sched
```

Bug reports where the same bug occurred

Fragments of code where the same bug was fixed

SHOWCASE

ADVANCED QUERY MODE

API

DATASETS

PUBLICATIONS

ABOUT

# BIANCA: Preventing Bug Insertion at Commit-Time Using Clone Detection

- BIANCA learns known defects by mining BUMPER-indexed systems.
- It intercepts developer's code and compares it to signatures of known defects.
- If a match exists, a flag is raised and a fix is proposed.

```
48 86 F7 0D 01 07 02 A0 82 24 0C 30 82 24 08 02
01 01 31 0B 30 09 06 05 2B 0E 03 02 1A 05 00 30
68 06 0A 2B 06 01 04 01 82 37 02 01 04 A0 5A 30
58 30 33 06 0A 2B 06 01 04 01 82 37 02 01 0F 30
25 03 01 00 A0 20 A2 1E 80 1C 00 3C 00 3C 00 3C
00 4F 00 62 00 73 00 6F 00 6C 00 65 00 74 00 65
00 3E 00 3E 00 3E 30 21 30 09 06 05 2B 0E 03 02
1A 05 00 04 14 DB F1 70 2C DC 6E EC 31 15 51 EB
DC 94 F4 26 FC A2 8F 0E 69 A0 82 1E E1 30 82 04
12 30 82 02 FA A0 03 02 01 02 02 0F 00 C1 00 8B
3C 3C 88 11 D1 3E F6 63 EC DF 40 30 0D 06 09 2A
86 48 86 F7 0D 01 01 04 05 00 30 70 31 2B 30 29
06 03 55 04 0B 13 22 43 6F 70 79 72 69 67 68 74
20 28 63 29 20 31 39 39 37 20 4D 69 63 72 6F 73
6F 66 74 20 43 6F 72 70 2E 31 1E 30 1C 06 03 55
04 0B 13 15 4D 69 63 72 6F 73 6F 66 74 20 43 6F
72 70 6F 72 61 74 69 6F 6E 31 21 30 1F 06 03 55
04 03 13 18 4D 69 63 72 6F 73 6F 66 74 20 52 6F
6F 74 20 41 75 74 68 6F 72 69 74 79 30 1E 17 0D
39 37 30 31 31 30 30 37 30 30 30 30 5A 17 0D 32
30 31 32 33 31 30 37 30 30 30 30 5A 30 70 31 2B
30 29 06 03 55 04 0B 13 22 43 6F 70 79 72 69 67
68 74 20 28 63 29 20 31 39 39 37 20 4D 69 63 72
6F 73 6F 66 74 20 43 6F 72 70 2E 31 1E 30 1C 06
03 55 04 0B 13 15 4D 69 63 72 6F 73 6F 66 74 20
43 6F 72 70 6F 72 61 74 69 6F 6E 31 21 30 1F 06
03 55 04 03 13 18 4D 69 63 72 6F 73 6F 66 74 20
52 6F 6F 74 20 41 75 74 68 6F 72 69 74 79 30 82
```

TABLE 3: BIANCA results in terms of organization, project name, a short description, number of class, number of commits, number of defect introducing commits, number of risky commit detected, precision (%), recall (%), F<sub>1</sub>-measure (%), the average similarity of first 3 and 5 proposed fixes with the actual fix and the average time difference between detected and original.

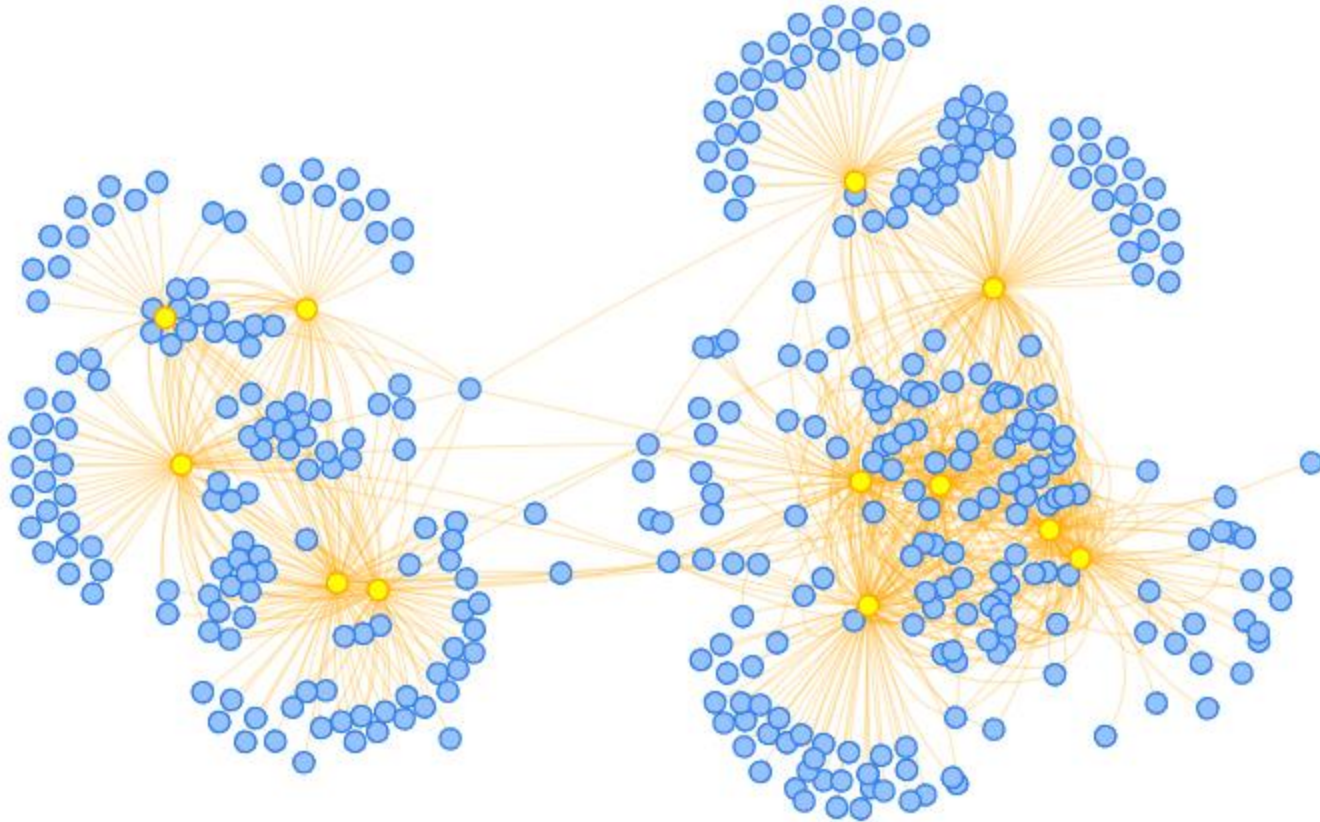
Organization	Project Name	Short Description	NoC	#Commits	Bug Introducing Commit	Detected	Precision	Recall	F <sub>1</sub>	Top 5 Fixes Similarity	Top 3 Fixes Similarity
Alibaba	druid	Database connection pool	3,309	4,775	1,260	787	88.44	62.46	73.21	39.97	46.69
	dubbo	RPC framework	1,715	1,836	119	61	96.72	51.26	67.01	60.01	57.14
	fastjson	JSON parser/generator	2,002	1,749	516	373	95.71	72.29	82.37	18.19	15.23
	Stream Process	Stream Process	1,492	215	24	21	90.48	87.50	88.96	22.38	30.48
Apache	hadoop	Distributed processing	9,108	14,154	3,678	851	86.84	23.14	36.54	38.94	47.68
	storm	Realtime system	2,209	7,208	951	444	86.26	46.69	60.58	53.03	61.10
Clojure	clojure	Programming language	335	2,996	596	46	86.96	7.72	14.18	53.61	59.52
Dropwizard	dropwizard	RESTful web services	964	3,809	581	179	96.65	30.81	46.72	47.54	53.56
	metrics	JVM metrics	335	1,948	331	129	95.35	38.97	55.33	22.53	31.82
Eclipse	che	Eclipse IDE	7,818	1,876	160	0	88.80	5.33	10.05	31.01	30.01
E											
F											
G											
G											
Ja											
In											
L											
N											
O											
Openarkun	zipkin	Distributed tracing system	371	1,777	119	73	87.87	41.90	50.31	32.76	31.70
Orfjackal	retrolambda	Backport of Java 8's lambda	171	447	97	35	94.29	36.08	52.19	34.69	42.06
Orient Technologie	orientdb	Multi-Model DBMS	2,907	13,907	7,441	2,894	86.77	38.89	53.71	62.20	70.00
Perwendel	spark	Sinatra for java	205	703	125	82	97.56	65.60	78.45	21.88	28.00
PrestoDB	presto	Distributed SQL query	4,381	8,065	2,112	991	90.62	46.92	61.83	23.34	20.64
Roboguice	roboquice	Google Guice on Android	1,193	1,053	229	70	91.43	30.57	45.82	53.81	56.55
Lombok	lombok	Additions to the Java language	1,146	1,872	560	212	91.98	37.86	53.64	58.94	57.49
Scribejava	scribejava	OAuth library	218	609	72	16	93.75	22.22	35.93	30.05	38.16
Square	dagger	Dependency injector	232	697	144	84	90.48	58.33	70.93	64.29	64.97
	javapoet	Java API	66	650	163	113	100.00	69.33	81.88	51.04	53.20
	okhttp	HTTP+HTTP/2 client	344	2,649	592	474	93.04	80.07	86.07	29.09	24.91
	okio	I/O API for Java	90	433	40	24	100.00	60.00	75.00	31.51	35.50
	otto	Guava-based event bus	84	201	15	15	93.33	100.00	96.55	54.11	49.94
	retrofit	Type-safe HTTP client	202	1,349	151	111	99.10	73.51	84.41	49.88	45.46
StephaneNicolas	robospice	Android library	461	865	113	39	87.18	34.51	49.45	60.90	65.04
ThinkAurelius	titan	Graph Database	2,015	4,434	1,634	527	90.13	32.25	47.51	48.64	50.59
Xetorthio	jedis	Redis client	203	1,370	295	226	92.04	76.61	83.62	25.69	29.45
Yahoo	antheion	Plugin for Apache Nutch	1,620	7	0	-	-	-	-	-	-
Zxing	zxing	1D/2D barcode image	3,030	3,253	791	123	94.31	15.55	26.70	29.35	37.96
<b>Total</b>			<b>96,003</b>	<b>165,912</b>	<b>41,225</b>	<b>15316</b>	<b>90.75</b>	<b>37.15</b>	<b>52.72</b>	<b>40.78</b>	<b>44.17</b>

- Subject systems: 42 open source projects
- Precision = 90% and Recall: 37%
- BIANCA fixes are accurate in 79% of the cases

# CLEVER: Combining Levels of Bug Prevention and Resolution Techniques

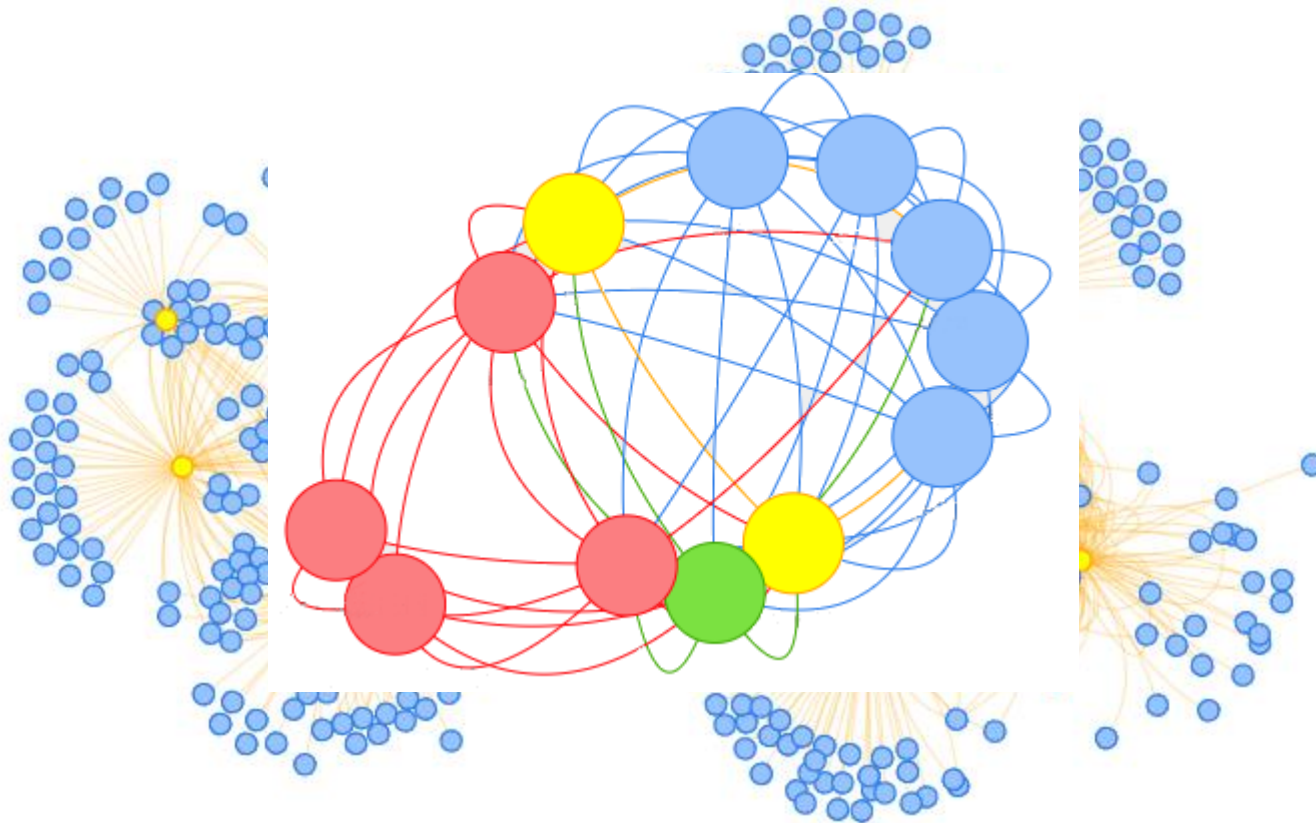
- Combines multiple features to determine the defect signatures
- Uses domain expertise to create clusters of projects for improved accuracy
- Uses better code matching techniques
- Is evaluated on 12 Ubisoft systems

# CLEVER Project Clustering



We can improve the detection accuracy if we search within inter-related projects

# CLEVER Project Clustering (Cont.)



We can improve the detection accuracy if we search within inter-related projects

# Evaluation of CLEVER at Ubisoft

- **Results:**
  - Subject systems: 12 Ubisoft systems
  - Precision = 79% and Recall = 65%
  - CLEVER recommends fixes in 67% of the cases
- **Impact on productivity:**
  - CommitAssistant (internal implementation of CLEVER) is designed to integrate well with developers' workflow
  - Ubisoft announced in a press release that CommitAssistant can cut the bug fixing time by 20%



# Software Development

A set of activities for creating a software system including requirements analysis, architectural design, detailed design, coding, testing, maintenance, integration, acceptance testing, etc.

# Software Operations

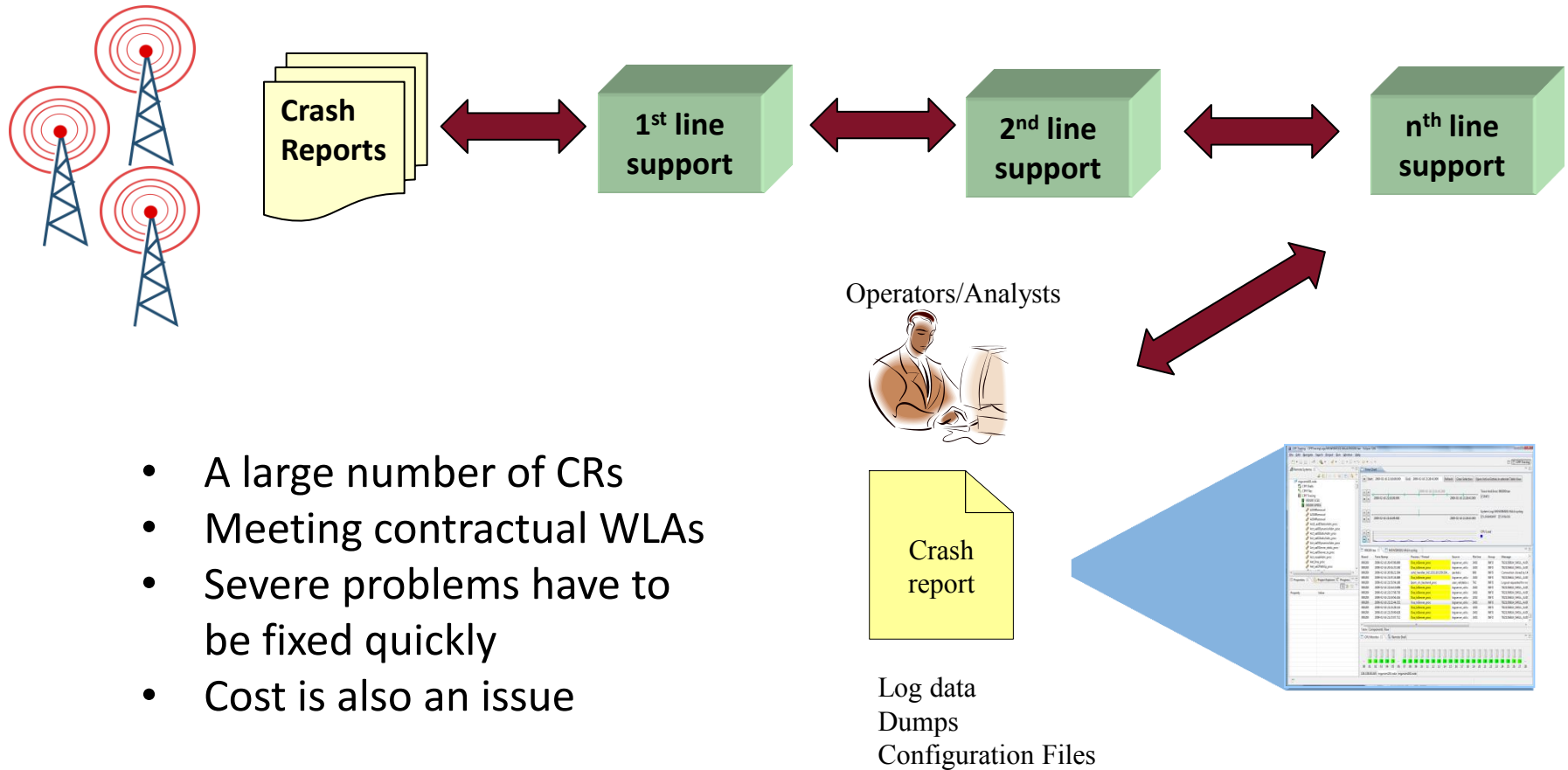
A set of activities for supporting end users of a software product in an operational environment. Typical activities include: installation, upgrade, monitoring, configuration, etc.

Source: Definitions adapted from ISO/IEC24748-1 2011, ISO/IEC15288 2008, ISO/IEC12207 2008).

# Software Operations

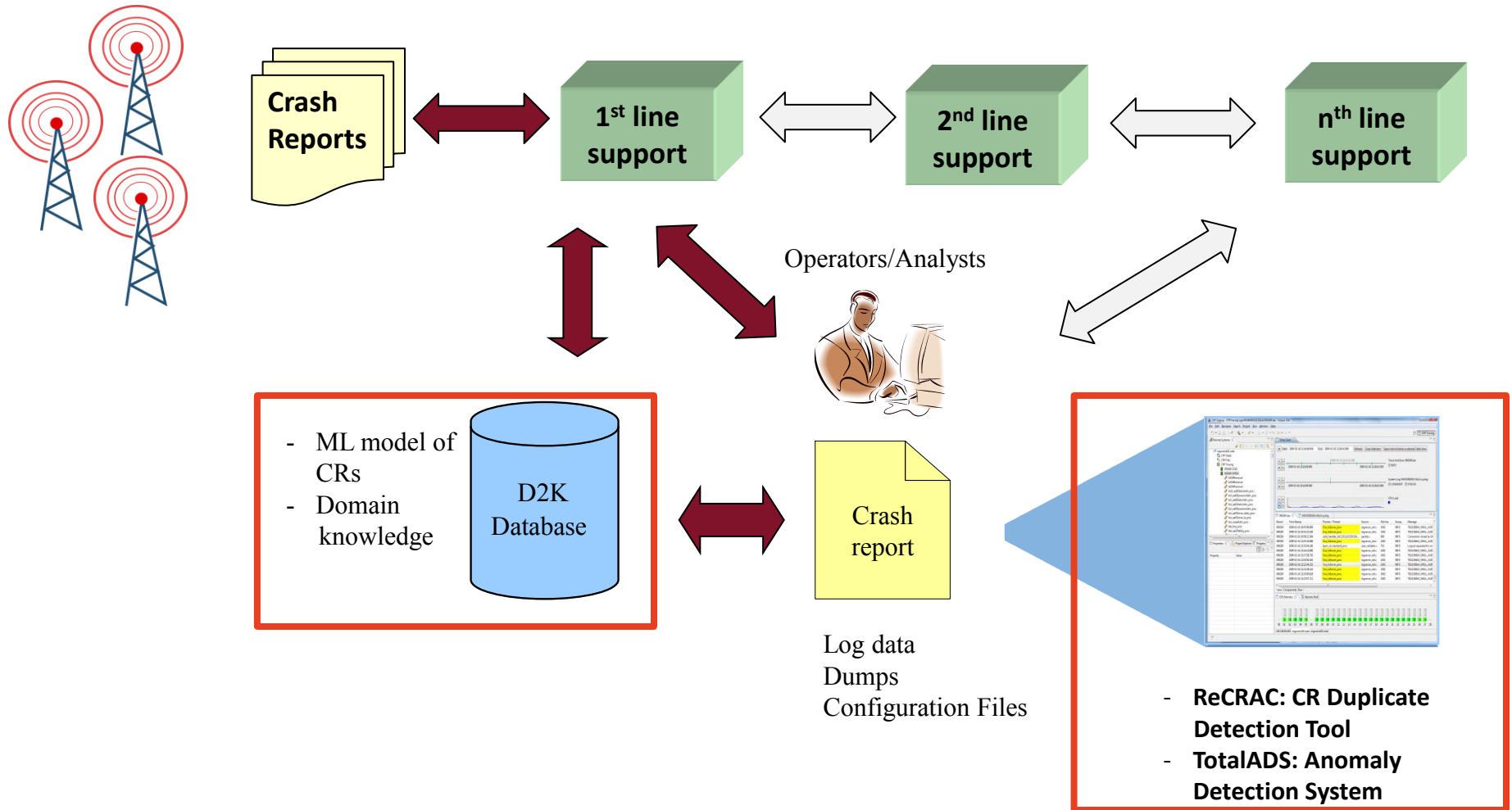
- D2K Project: From Data To Knowledge for Better System Maintenance
- Collaborators: Ericsson, NSERC, and MITACS
- Objectives:
  - Improve the crash report (CR) handling process
  - Investigate automated solutions
  - Provide analysis capabilities to operators
  - Provide data governance solutions

# Problem



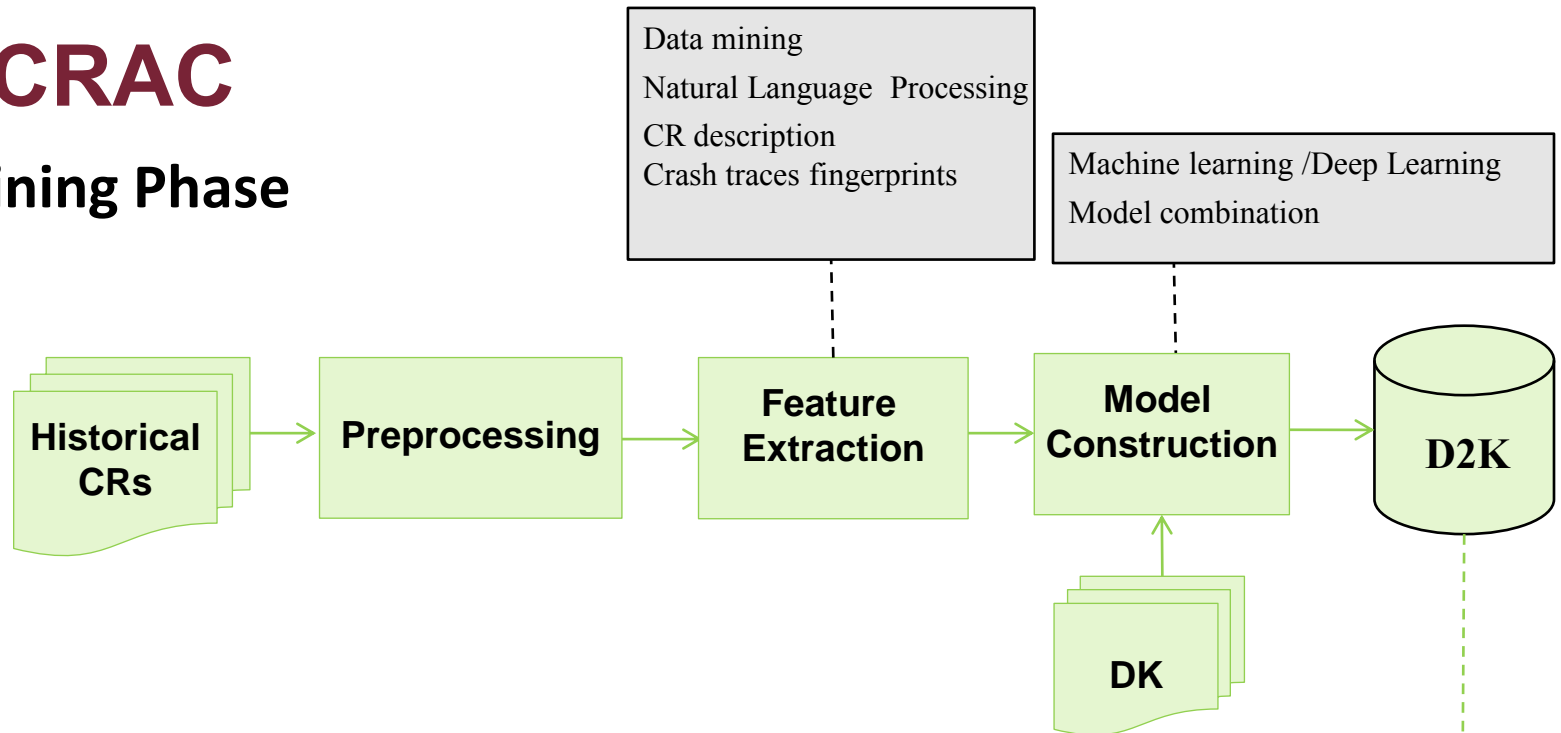
- A large number of CRs
- Meeting contractual WLAs
- Severe problems have to be fixed quickly
- Cost is also an issue

# Solution

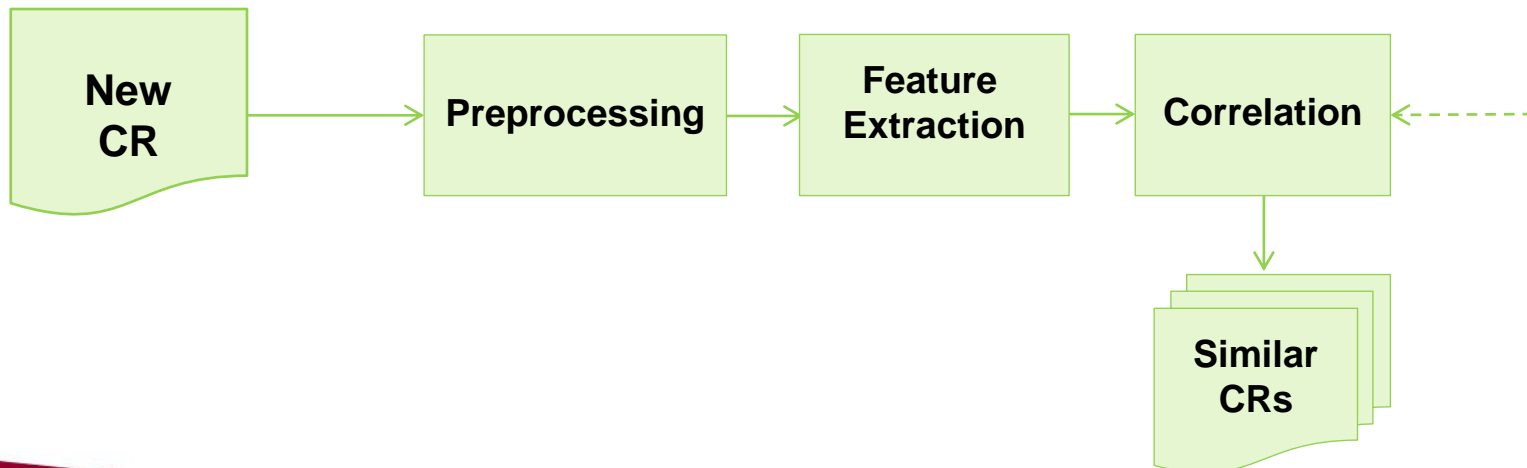


# ReCRAC

## Training Phase



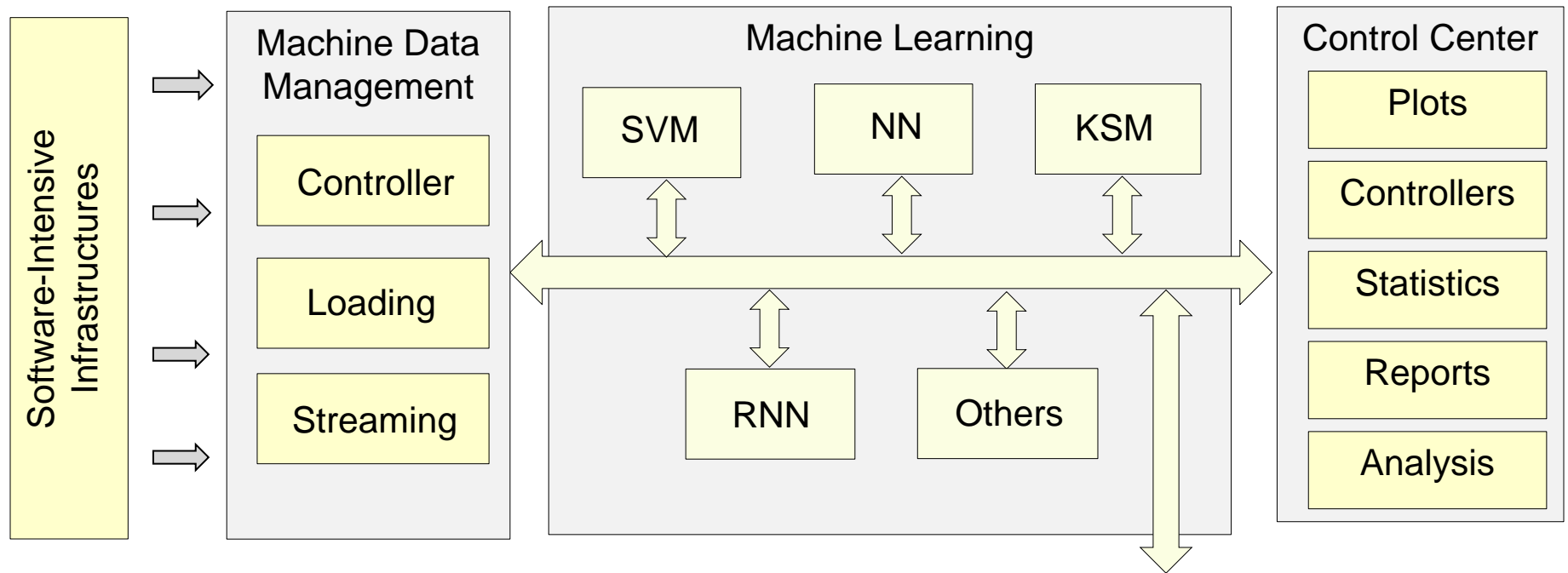
## Testing Phase (ReCRAC in operation)



# TotalADS: Total Anomaly Detection System

- Developed in an NSERC project with Defence R&D Canada and Ericsson
- Objectives:
  - Detection of abnormal behavior in computer hosts through the analysis of machine data
  - Combination of multiple machine learning techniques
  - Leverage of data abstraction, model combination, adaptive learning, and online learning
  - Tool development and integration

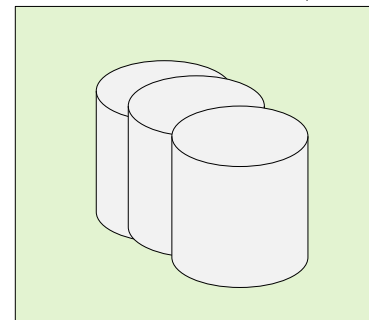
# TotalADS: Total Anomaly Detection System Architecture



Data Centers  
Radio Stations  
Smart Grids  
IoT Devices



IBM CASCON 2014  
PEOPLE'S CHOICE  
AWARD

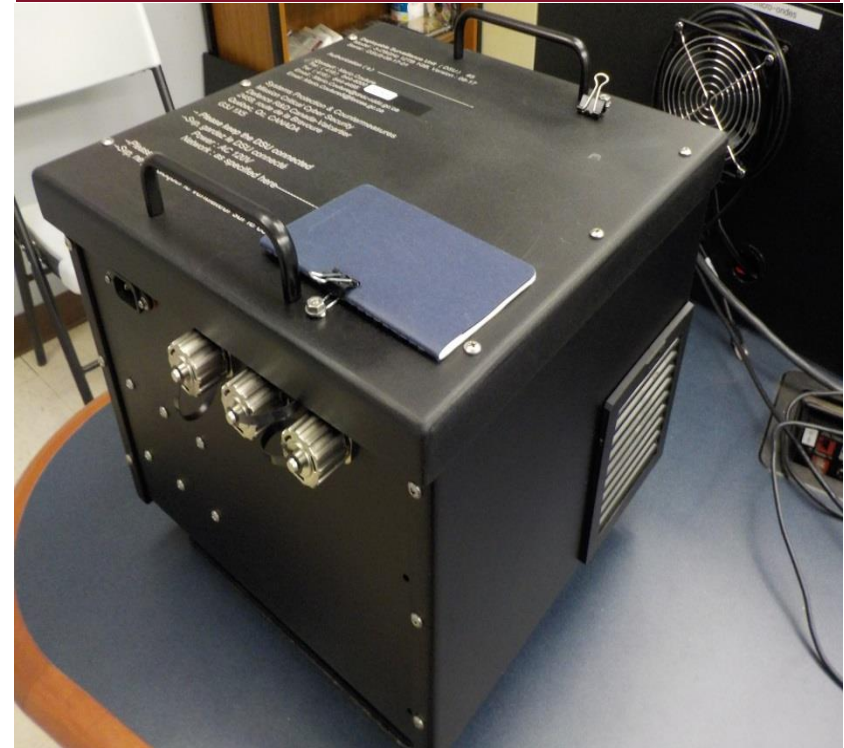


# TotalADS and Deployable System Units (DSUs)

## The 4<sup>th</sup> DSU prototype (PoC1)



## The 5<sup>th</sup> DSU prototype (PoC2)



Six Jetway industrial Mini-ITX computers + one manageable GB switch + six 4-TB hard disks  
(Intel's Haswell Core i7-4770TE 2.3 GHz processor, 8 GiB DDR3, 6 GB/sec mSATA, dual LAN)  
(The whole DSU needs less than 350 watts when used at full capacity)

(Next technology to be considered: the new NVIDIA Jetson TX-2 AI computing board)



# Some thoughts on the use of DL/ML in SW Development and Operations



**Powerful tool suite**



**Healthy analytics**



**Context matters**



**Domain expertise**



**Education**



**Impact on society**

**THANK YOU!**



CONCORDIA.CA