

How Should I Fix This Bug? - Investigating Text Analytics and Social Network Analysis Based Approaches to Support a Practitioner in Bug Fixing

Ashish Sureka [ashish@iiitd.ac.in]

Adjunct Faculty, IIT-Delhi and Visiting Researcher, Siemens Research
Founder, Software Analytics Research Lab (SARL)



SoftwareAnalytics.in
Mining Software Archives Research at IIT - Delhi

Principal Investigator and Project Details



Prof. Ashish Sureka

Adjunct Professor at IIIT-Delhi

Visiting Researcher at Siemens Corporate Research and Technology

Independent Consultant, Educator and Researcher (Software Analytics)

Website: <http://www.software-analytics.in/>

Ref. No.	SR/FTP/ETA- 106/2010
Project Duration	3 Years
Date of Sanction	15/05/2012
Date of Start	01/07/2012
Total Sanctioned Cost	Rs. 342000
Funds Released so far	Rs. 205000
Expenditure incurred as on 13/12/2014	Rs. 160000



Presentation Outline

- Original objectives and starting date
- Original schedule & progress, achievements
- Reason for deviation, if any
- Any mid-course changes and why
- New knowledge acquired
- Expense budget v/s actual
- Timely completion and within budget?
- Publications, Thesis and Graduated Students



Research Motivation

- ✓ Software development is human-intensive task and is prone to error
- ✓ Software needs to be maintained after delivery/release (*maintenance is inevitable !*)
- ✓ Maintenance = **correcting faults**, improving performance, adapting to changing environments
- ✓ Software maintenance in **large & complex** setting is challenging and expensive
- ✓ Studies shows : maintenance = more than 50% financial resources of software lifetime cost
- ✓ Corrective maintenance and defect removal = 15-60% within maintenance activities
- ✓ The *focus* of this (How Should I Fix This Bug) work: **Bug-Fixing after Delivery/Release**



Original Objectives and Starting Date

Date of Sanction	15/05/2012
Date of Start	01/07/2012

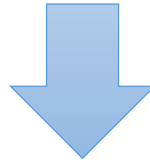
Research Aim

- ✓ Study the **post-delivery processes** and activities in context to maintenance
- ✓ Identify technical **challenges encountered by practitioners** in corrective maintenance
- ✓ Investigate techniques to **assist a practitioner** in corrective maintenance tasks



Research Aim:

- ✓ Study the **post-delivery processes** and activities in context to maintenance
- ✓ Identify technical **challenges encountered by practitioners** in corrective maintenance
- ✓ Investigate techniques to **assist a practitioner** in corrective maintenance tasks



SURVEYS

Open Source Projects (OSS)

- ✓ Google Chromium
- ✓ Google Android
- ✓ Mozilla Firefox
- ✓ Eclipse
- ✓ Open Office

Close Source (CSS) and Commercial

- ✓ Infosys
- ✓ TCS
- ✓ Wipro
- ✓ ADOBE
- ✓ A 200 employee, mid-size company

- Core Developers, Testers, Bug Fixers, Bug reporters, Bug Traigers, QA (Quality Assurance) Managers
- Survey results are presented in papers
- What is the impact of impact of SE research on practice and practitioners-researchers/academics collaboration has been a major (even provocative) topic of discussion

Research Framework and Architecture



Tester



Triager



Developer



Manager

Duplicate bug report detection
 Bug report allocation and expertise modeling
 Identifying location of a bug in source code
 Identifying defect-prone & attack-prone modules

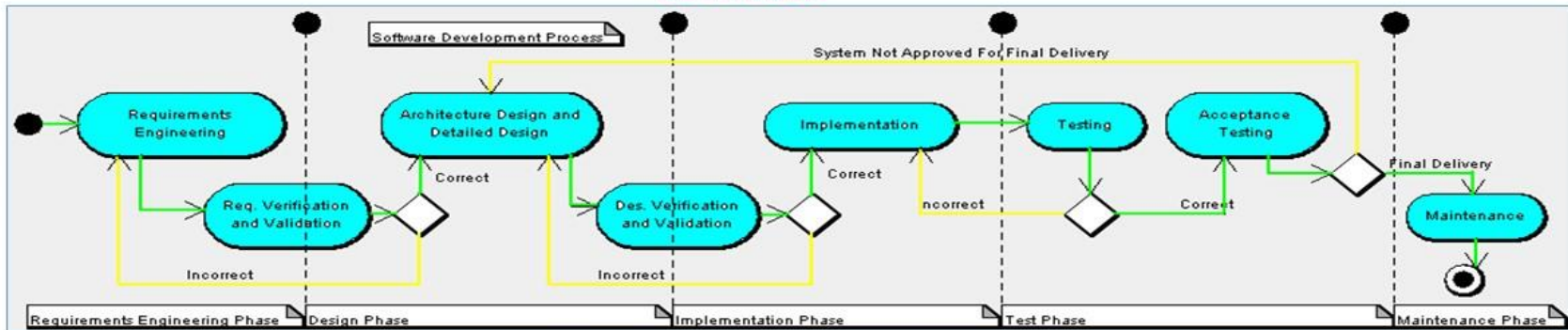
Machine Learning

Social Network Analysis

Text Analytics

Data Mining

TECHNIQUES



SOFTWARE DEVELOPMENT PROCESS



ISSUE TRACKING SYSTEMS



COMMUNITY BASED Q&A



PEER CODE REVIEW SYSTEM



VERSION CONTROL SYSTEM

SOFTWARE REPOSITORIES



Process Mining Software Repositories from Student Projects in an Undergraduate Software Engineering Course

36th International Conference on Software Engineering, SE Education and Training Track (ICSE SEET 2014)

Software Engineering is a practice oriented and applied discipline

Learning objectives

- Learn fundamental technical concepts in SE courses.
- Learn software development processes, teamwork and project management.
- Exposure to popular SE tools.

- ✓ Course Instructor can easily assess and provide feedback on product
- ✓ *Providing feedback on process is not straightforward*

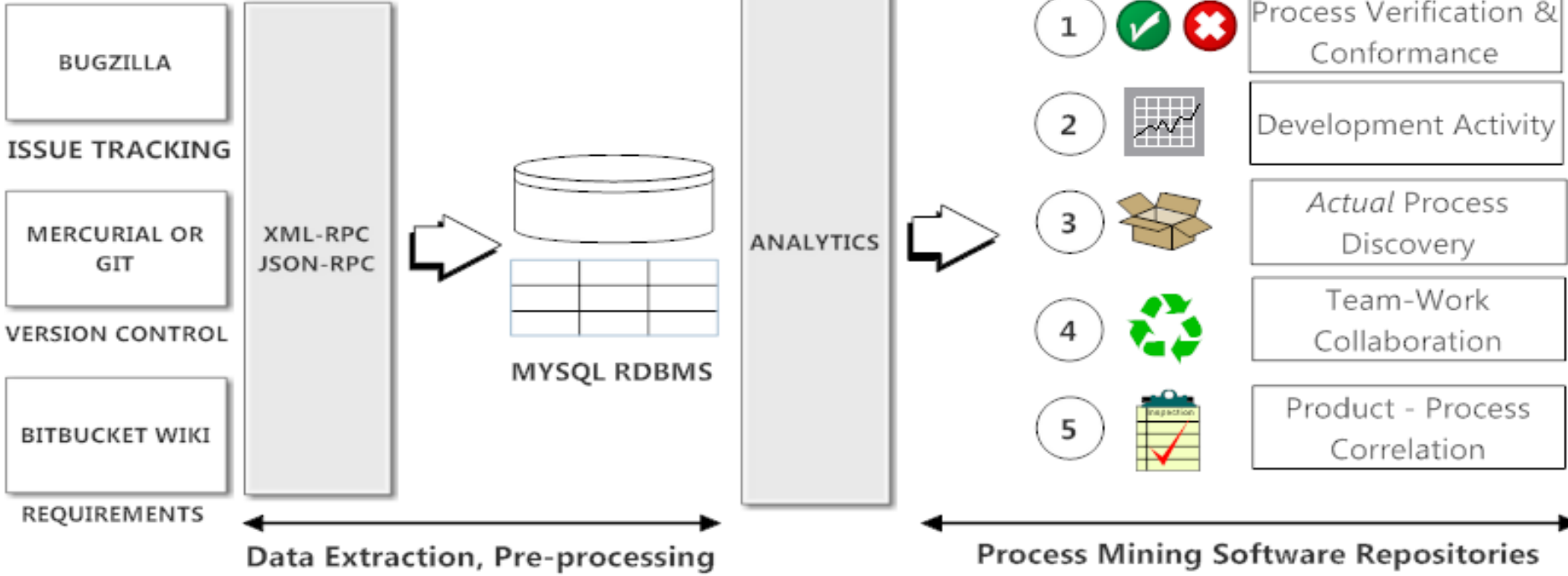
AIM

- Gain visibility and insights on software development processes followed by student teams

- Provide appropriate feedback on process improvement



RESEARCH FRAMEWORK



Progress and Achievements



No deviation and no mid-course changes

Timely completion and within budget



Published several papers. Results dissemination

Published in Rank A+ [Exceptional] conferences

Presented tutorials in good conferences



Graduated several students

PhD and MTech Thesis/Dissertations



Graduated Students [MTech]

06/Jan/2014	Megha Mittal	Mimansa: Process Mining Software Repositories from Student Projects in an Undergraduate Software Engineering Course
18/Feb/2014	Ritika Jain	Sanayojan: A Framework for Traceability Link Recovery between Use-Cases in Software Requirement Specification and Regulatory Documents
13/Oct/2014	Yash Lamba	Pravaaha: Mining Android Applications for Discovering API Call Usage Patterns and Trends
13/Oct/2014	Manisha Khattar	Sarathi: Characterization Study on Regression Bugs and Identification of Regression Bug Inducing Changes: A Case-Study on Google Chromium Project
13/Feb/2015	Nisha Gupta	Pariket: A Framework for Mining Business Process Logs for Root Cause Analysis of Anomalous Incidents

Graduated Students [PhD]

20/Dec/2014	Denzil Correa	Content Quality in Online Services - Analysis, Detection and Enhancement
--------------------	---------------	--



Research Paper Publications

1	ISEC 2015	Software Process Intelligence: Mining Software Process Data for Extracting Actionable Information
2	ISEC 2015	Pravaaha: Mining Android Applications for Discovering API Call Usage Patterns and Trends
3	ISEC 2015	Sarathi: Characterization Study on Regression Bugs and Identification of Regression Bug Inducing Changes: A Case-Study on Google Chromium Project
4	COMAD 2014	Kashvi: A Framework for Software Process Intelligence
5	APSEC 2014	MiQs: Characterization and Prediction of Migrated Questions on StackExchange
6	MUD 2014	Mining Peer Code Review System for Computing Effort and Contribution Metrics for Patch Reviewers
7	RAISE 2014	Sanayojan: A Framework for Traceability Link Recovery between Use-Cases in Software Requirement Specification and Regulatory Documents



Research Paper Publications

8	RAISE 2014	Mining Issue Tracking Systems Using Topic Models for Trend Analysis, Corpus Exploration and Understanding Evolution
9	ICSE SEET 2014	Process Mining Software Repositories from Student Projects in an Undergraduate Software Engineering Course
10	WWW 2014	Chaff from the Wheat: Characterization and Modeling of Deleted Questions on Stack Overflow
11	APSEC 2013	Samekana: A Browser Extension for Including Relevant Web Links in Issue Tracking System Discussion Forum
12	COSN 2013	Fit or Unfit: Analysis and Prediction of Closed Questions on Stack Overflow
13	ASWEC 2013	Integrating Issue Tracking Systems with Community-Based Question and Answering Websites
14	APSEC 2012	Empirical Analysis of Bug Report Types: A Case-study of Google Chrome Browser Project
15	TOOLS 2012	Learning to Classify Bug Reports into Components



Budget and Finances

Budget Heads	Approved amount (First Installment)	Expenditure Incurred Till Date	Balance left unutilized
Equipments	65,000	59035	5,965
Consumables	15,163	0	15,163
Contingencies	30,325	30325	0
Other/Analytical	35,380	26329	9,051
Overheads	28,808	28808	0
Travelling	30,324	15000	15,324
Total	2,05,000	1,59,497	45,503

