

Mining-Ansätze für Software Repositories

Nataliia Lashchuk

Seminar „Beiträge zum Software Engineering“

Übersicht

Einführung

Mining Software Repositories (MSR)

Repositories

Fokus & Umfang

Werkzeuge

Artikel I

Ziel und Methodik

Artikel II

Ziel und Methodik

Zusammenfassung

MSR Definition

„analyzing the rich data available in software repositories to uncover interesting and actionable information about software systems and projects“

[2015.msrconf.org]

Repositories

- Versionsverwaltungssysteme (Git, SVN)

Allg. Struktur eines Log Entry

Log Entry
Commit Message
Revision
Committer's username
Commit date
List of affected files
Commit type

Repositories

- Issue-Tracking-Systeme (Bugzilla, JIRA)

Allg. Struktur eines Fehlerberichts

Bug Report	
ID	Priority
Assignee	Product
Buggy Component	Summary
Date	Last changed time
Bug reporter	Remaining Time
Deadline	Target Milestone

Repositories

- Ablagen für Dokumentation (Wiki)
- Kommunikationsarchive (Mails, Mailingslisten, Foren)

MSR Themen

- Vorhersage der Softwarequalität
 - Fehlervorhersage
 - Wieviele Fehler werden in einem Modul erwartet?
 - In welchen Module werden diese auftreten?
- Analyse von Änderungen und Korrekturen
- Entwicklungsaufwand
- Verständnis der Zusammenarbeit in Teams
- Klassifizierung und Lokalisierung von Fehlern bzw. Änderungen

MSR Werkzeuge

➤ DynaMine

Kodierungsmuster

➤ CVSAAnaly

Log Parser für Quellcode Repository

➤ Bicho

Issue Parser

➤ Relink

Verhältnisse: Quellcode Repository <-> Bugtracker

Referenzarbeiten

I

Process Mining Multiple Repositories for Software Defect Resolution from Control and Organizational perspective

Monika Gupta, Ashish Sureka, Srinivas Padmanabhuni
Hyderabad, India, 2014.

II

Using Social Network Analysis for Mining Collaboration Data in a Defect Tracking System for Risk and Vulnerability Analysis

Ashish Sureka, Atul Goyal, Ayushi Rastogi
New Delhi, India, 2011.

I: Ziel

- Untersuchung der Ansätze zum Extrahieren von Textdaten anhand:
 - Beschreibung eines Fehlerberichts und dessen Kommentare;
 - Beschreibung der Codereview;
 - Commit-Nachrichten
- Entwicklung eines Event Logs zum Beobachten des Fehlerverlaufs
 - Erstellen eines Fehlerbericht bis zum Schliessen

I: Untersuchungsdomäne

Chromium Projekt

35035 geschlossenen Fehler

- issue tracking system (Google ITS)
- peer code review System (Rietveld)
- version control system (Subversion)

I: Forschungsmethodik

➤ Datenextraktion und Intergration

JSON-RPC

XML-RPC

➤ Datentransformation

MySQL

➤ Data Mining

I: Datenextraktion und Integration

Issue 88294: Default printing settings are always "two s
1 person starred this issue and may be notified of changes.

Status: Fixed
Owner: kmadhu_@chromium.org
Closed: Jul 2011
Type-Bug
Pri-2
OS-Windows
Cr-Internals
M-14

Reported by fam_@live.nl Jul 2, 2011
Chrome Version : 14.0.803.0
OS Version: Windows 7
URLs (if applicable) : print preview pa

What steps will reproduce the problem?
1. My printer (HP Deskjet 3600) doesn't this printer in print preview.
2. It does support color prints. Still,

[#2 kmadhu_@chromium.org](#)

The following revision refers to this bug:
<http://exp.chromium.org/viewvc/chrome?view=rev&revision=92154>

[r92154](#) | kmadhusu@chromium.org | Tue Jul 12 09:59:14 PDT 2011

Changed paths:
M <http://exp.chromium.org/viewvc/chrome/trunk/src/chrome/browser/printpreview.cc>
A <http://exp.chromium.org/viewvc/chrome/trunk/src/chrome/browser/printpreview.cc>

PrintPreview: [WIN] Fix the default duplex print setting.

BUG=88294
TEST=Please refer to bug report.

Review URL: <http://codereview.chromium.org/7285039>

ITS BUG ID

PCR ISSUE ID

ISSUE TRACKING SYSTEM

Issue 7285039: PrintPreview: [WIN] Fix the default duplex print setting. (Closed)

Can't Edit
Can't Publish+Mail
[Start Review](#)

Created: 2 years, 5 months ago by kmadhusu
Modified: 2 years, 5 months ago
Reviewers: Lei Zhang, [[hide the power \(commit-bot\)](#)]
CC: chromium-reviews_chromium.org

Description
PrintPreview: [WIN] Fix the default duplex print setting.

BUG=88294
TEST=Please refer to bug report.

Committed: <http://exp.chromium.org/viewvc/chrome?view=rev&revision=92154>

Patch Set 1
Total comments: 2

Patch Set 2 : Fixed nit

PCR ISSUE ID

ITS BUG ID

Revision 92154

Jump to revision:

92154

Go

Author:

kmadhusu@chromium.org

Date:

Tue Jul 12 12:55:14 2011 UTC (2 years, 5 months ago)

Changed paths: 1

Log Message:

PrintPreview: [WIN] Fix the default duplex print setting.

BUG=88294

TEST=Please refer to bug report.

Review URL: <http://codereview.chromium.org/7285039>

PEER CODE
REVIEW SYSTEM

VCS REVISION ID

VERSION CONTROL
SYSTEM

PCR ISSUE ID

I: Forschungsmethodik

➤ Datenextraktion und Intergration

JSON-RPC

XML-RPC

➤ **Datentransformation**

MySQL

➤ Data Mining

I: Tätigkeiten bei einem Fehler mit Prozessauswirkung

Information System	Activity	Description	Role
Issue Tracking System	I_Creation	Issue reported in ITS	Bug Reporter
	I_Open	Open bug status label	Bug Owner, Triager, QA Manager
	I_Fixed	Resolved as Fixed	Bug Owner
	I_Invalid	Illegible, spam, etc.	Bug Owner, Triager, QA Manager
	I_Duplicate	Similar to other issue	Bug Owner, Triager, QA Manager
	I_WontFix	Can't repro, Working as intended, Obsolete	Bug Owner, Triager, QA Manager
	I_Verified	Resolution verified	QA Manager
	I_Closed	ITS progress ends	Bug Owner
Code Review System (suffix: seq. no.)	C_Creation	Initial patch reported in PCR	Patch Author
	P_Creation	Subsequent patch submit with corrections	Patch Author
	C_Reviewed	Code review process ends	Patch Reviewer
Version Control System	V_Commit	Code change committed	Patch Committer

I: Forschungsmethodik

➤ Datenextraktion und Intergration

JSON-RPC

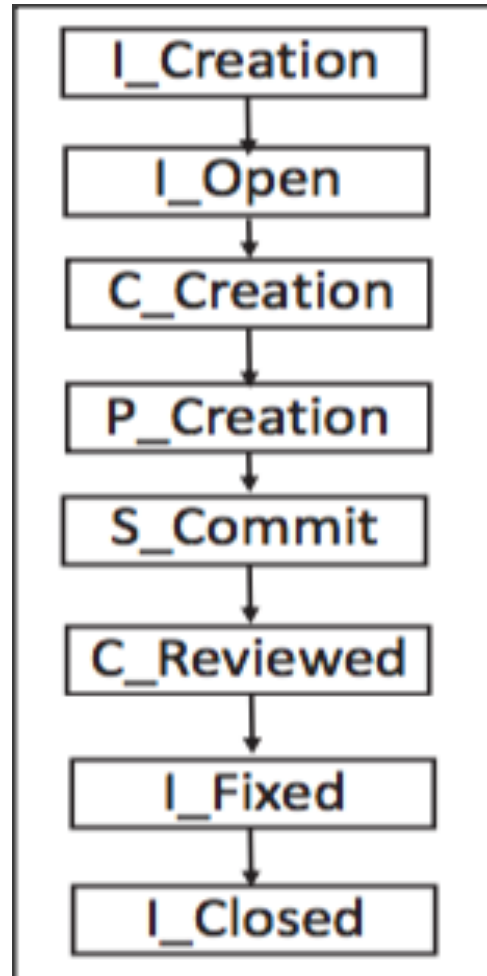
XML-RPC

➤ Datentransformation

MySQL

➤ **Data Mining**

I: Prozess-Modell zur Lösung eines Fehler



II: Ziel

- Analyse von Fehlerberichten
- Untersuchung der Techniken für Mining Bug-Tracking-Systeme, welche die Risiken und Schwachstellen abdecken können

II: Untersuchungsdomäne

- Defect Tracking System (Bugzilla)
- Projekt Mozilla Firefox

II: Forschungsmethodik

➤ Untersuchung der Kommunikation der Entwickler in
Kommentare

Kooperationsnetzwerk

core

peripheral

II: Forschungsmethodik

➤ Untersuchung der Beziehung zwischen Schweregrad eines Fehlers und Entwickler

Welche Entwickler arbeiten mit kritischen Fehlern?

Welche Entwickler arbeiten mit einer Vielzahl von Schweregraden der Fehler?

Gibt es Entwickler, die nur mit einem gewissen Schweregrad der Fehler arbeiten?

II: Forschungsmethodik

➤ Untersuchung der Autorität der Entwickler

Welche Entwickler haben eine Maximalzahl von Verbindungen?

Brücken

II: Forschungsmethodik

- Untersuchung der Beziehungen zwischen
Komponenten und Schweregrade der Fehler
 - Fehleranfälligkeit der Komponenten
 - Verhältnis der Entwickler zu defektanfälligen Komponenten

II: Forschungsmethodik

➤ Untersuchung der Untergruppen

Gib es Entwickler-Gruppen, die häufig kooperieren?

Position der Entwickler in der Gemeinschaft

II: Risiko- und Verletzlichkeit-Perspektive

- Verständnisse über die Anzahl von Core-Entwickler
- Verständnisse über die Anzahl von Entwickler, die an kritischen Fehlern arbeiten
- Verständnisse über die Entwickler, welche mit den bestimmten Komponenten arbeiten

Zusammenfassung

- Domäne (I, II): Defekte
- Großer manueller Aufwand
- Verschiedene Perspektive
 - I: Prozess und Organisation
 - II: Risiko
- Repositories:
 - I: ITS, PCR, VCS
 - II: ITS
- Fehlerberichte
 - I: geschlossene
 - II: alle, die bereits kommentiert wurden

Vielen Dank!