Code Inspection and Code Review
One-Slide Summary

• In a **code review**, another developer examines your proposed change and explanation, offers feedback and decides whether to accept it. Modern code reviews have significant **tool support**.

• In a **(formal) code inspection**, a team of developers meet to examine existing code, following a process to understand it and spot issues.

• Both of these static quality assurance approaches have **costs** and **benefits**.
The Story So Far ...

- Quality assurance is critical to software engineering.

- Testing is the most common dynamic ("run the program") approach to QA.
  - We can generate some test inputs and oracles, but testing remains very expensive.

- What about static ("look at the program") approaches to QA?
Intuition

“Given enough eyeballs, all bugs are shallow.”

- Linus Law

“Have peers, rather than customers, find defects.”

- Karl Wiegers
Example of Both: Twilight

The Cullens and the Hales sat at the same table as always, not eating, talking only among themselves. None of them, especially Edward, glanced my way anymore.

One person cannot do nothing more than other people who are also doing nothing.

None of them bought an apple, especially Edward.

I went to go try to watch TV while I waited.

Not only did Meyer make the sentence far more complicated than necessary, but she also made the act of watching television complicated.

Our heroine is so insecure that she actually doubts her ability to succeed at watching television by herself.

Our Bodies, Ourselves: *A Mystery*

By Isabella M. Swan

Yes, I wanted to say. Anything. But I couldn’t find my lips.

I tried to obey, though I couldn’t quite locate my lungs.
Why not simply test?

- Faults can mask other faults at runtime
- Only completed implementations can be tested (esp. scalability, performance)
- Many quality attributes (eg., security, compliance, maintainability) are hard to test
- Non-code artifacts (e.g., design documents) cannot be tested
A Second Pair of Eyes

- Different background, different experience
- No preconceived idea of correctness
- Not biased by “what was intended”
- “Breadth of experience in an individual is essential to creativity and hence to good engineering. ... Collective diversity, or diversity of the group - the kind of diversity that people usually talk about - is just as essential to good engineering as individual diversity. ... Those differences in experience are the "gene pool" from which creativity springs.”

- Bill Wulf, NAE President
What To Examine

- **Code Inspection**: Examine Whole Program
  - Expensive if the program changes
  - Good if a new concern arises

- **Code Review**: Examine Each Change
  - Inductive Argument:
    - $V_0$ is good, $V_n$ is good $\rightarrow$ $V_{n+1}$ is good
  - Bad if the definition of “good” changes
Code Inspection Example:
It's A Bug Hunt!

```c
year = ORIGINYEAR; /* = 1980 */
while (days > 365) {
    if (IsLeapYear(year)) {
        if (days > 366) {
            days -= 366;
            year += 1;
        }
    } else {
        days -= 365;
        year += 1;
    }
}
```
Leap-year glitch freezes Zune MP3 players

By Brandon Griggs
CNN

(CNN) -- A leap-year related glitch caused thousands of Zune MP3 players to simultaneously stop working late Tuesday and early Wednesday, Microsoft said on the product's Web site.

The problem should resolve itself after 7 a.m. ET Thursday, Matt Akers of the Zune Product Team wrote on Zune.net. "A bug in the internal clock driver related to the way the device handles a leap year" is to blame, he said.

The issue was limited to older Zune 30GB models, the Web site said.

The Zune support page says users should allow the internal battery to fully drain. Then, after noon GMT on January 1, 2009 (7 a.m. ET), users should recharge by connecting the Zune to a computer or AC power.

Internet message boards were flooded with complaints about Zunes freezing, prompting Y2K-like speculation about end-of-year hardware or software problems.

"It seems that every Zune on the planet has just frozen up and will not work," posted a Mountain Home, Idaho, user on CNN's iReport.com. "I have 3 and they all in the same night stopped working."
Code Review

- What is code review?
- What is different between code review and code inspection in practice?
GitHub

- **Pull requests** let you tell others about changes you've pushed to a repository on GitHub. Once a pull request is opened, you can discuss and review the potential changes with collaborators and add follow-up commits before the changes are merged into the repository.

- Other contributors can **review** your proposed changes, add review comments, contribute to the pull request discussion, and even add commits to the pull request.
Refactorings #28

Merged joliebig merged 17 commits into liveness from CallGraph 9 months ago

ckaestne commented on Jan 29

@joliebig
Please have a look whether you agree with these refactorings in CRewrite

key changes: Moved ASTNavigation and related classes and turned EnforceTreeHelper into an object

ckaestne added some commits on Jan 29

- remove obsolete test cases
- refactoring: move AST helper classes to CRewrite package where it is ...
- improve readability of test code
- removed unused fields

ckaestne commented on Jan 29

Can one of the admins verify this patch?
Microsoft
(Visual Studio, CodeFlow, etc.)

• Before you check in your code, you can use Visual Studio to ask someone else from your team to review it. Your request will show up in his team explorer, in the my work page.

• (Are you using Git to share your code? If so, then use a pull request.)
Dev #1 - Request Review
Dev #1 - Submit Request to Dev #2
Dev #2 - See and Accept Request

Hello World border color
Requested by Jamal Hartnett.

You can Accept or Decline to let the requestor know whether you will do the code review.

Reviewers (2)
Add Reviewer

- Johnnie McLeod - Requested
- Raisa Pokrovskaya - Accepted

Related Work Items
Dev #2 - View Details
Dev #2 - Suggest Improvements

Use #FF8C00 instead.
Google's Code Review Policy

- All change lists (“CLs”) must be reviewed. Period.
- Any CL can be reviewed by any engineer at Google.
- Each directory has a list of owners. At least one reviewer or the author must be an owner for each file that was touched in the commit. If the author is not in the owners file, the reviewer is expected to pay extra attention to how the code fits in to the overall codebase.
- One can enforce that any CLs to that directory are CC’d to a team mailing list.
- Reviews are conducted either by email, or using a web interface called Mondrian.
- In general, the review must have a positive outcome before the change can be submitted (enforced by perforce hooks). However, if the author of the changelist meets the readability and owners checks, they can submit the change “To Be Reviewed”, and have a post-hoc review. There is a process which will harass reviewers with very annoying emails if they do not promptly review the change.
Google, Facebook

• “In broad strokes, code review processes in Google and Facebook are similar. In both companies it is practically **required that every change to production code is reviewed** by at least one team member.

• Google has this readability process where you need to earn a privilege to commit in a given programming language. Readability is literally a badge on your profile that the code review system checks to see if you can commit the code yourself or you need to ask for an extra review for the compliance with company-wide language style guides.”

• Marcin Wyszynski 2017, worked at both companies
Tools

• Google uses Mondrian, an in-house tool
  • One of its authors later made https://www.gerritcodereview.com/
  • Reportedly, one of its authors later made https://reviewable.io/
  • Those give a taste of what it is like

• Facebook uses Phabricator
  • Developed in-house, later open-sourced https://www.phacility.com/
Fix daemon issues caused by Ubuntu's surprising intermediary shell  

Closed

Author: epriestley

Reviewers: rm, aran, tuomaspelkonen, jungesean, terabyte, punoet

CCs: aran, epriestley, rm, joleveley, hugobaraura, feynman, blii, ramk, w31rd0, dleyanlin, tailgasck, jiangzhongbo, tomlinsonryan, forestchu12, daviddeuler, abekkine, punoet, zaskary, lasseesperhoit, suwandi.cahyadi, lanceot, yao, ncu, rafatulta, jacob-zhoupeng, xiaoping, andreibelyaev, ganesanarunkumar, thangsp, jamesyu, googleyufe, demo, xiaoboz, alpha, jacobcy, michaelqyu, sowodyx, yool.annram, paprotnik123

Lint: List OK

Unit: No Unit Test Coverage

Commits: rPHIL7212040c896: Fix daemon issues caused by Ubuntu's surprising intermediary shell

Branch: master

Arcanist Project: libphull

Apply Patch: arc patch D212

Tokens: ☕️

epriestley summarized this revision.

On OSX and other Linuxli, proc_open("/exec_daemon ...") opens a PHP process; on Ubuntu it opens a "sh -c" process which opens a PHP process. The existence of this surprising shell made everything stop working.

Use 'exec' to replace the shell with the PHP process.

epriestley explained the test plan for this revision.

Ran daemons on OSX and Ubuntu. behavior seems okay in all cases.

Keep in mind I have absolutely no idea how Linux works so this probably breaks the world. (cc: simpkins)

epriestley commented on this revision.

See T298 for context.

rm accepted this revision.

Nice sleuthing

aran commented on this revision.

Hmm, I wonder what problem Ubuntu was solving by making that decision. Can it be configured? Also, is this the only callsite that will ever need this hack?
Code
Review
Integration
Example

(MediaWiki)
Do Code Reviews Work?

Expectations, Outcomes, and Challenges Of Modern Code Review

Alberto Bacchelli
REVEAL @ Faculty of Informatics
University of Lugano, Switzerland
alberto.bacchelli@usi.ch

Christian Bird
Microsoft Research
Redmond, Washington, USA
cbird@microsoft.com

Abstract—Code review is a common software engineering practice employed both in open source and industrial contexts. Review today is less formal and more "lightweight" than the code inspections performed and studied in the 70s and 80s. We empirically explore the motivations, challenges, and outcomes of tool-based code reviews. We observed, interviewed, and surveyed developers and managers and manually classified hundreds of review comments across diverse teams at Microsoft. Our study reveals that while finding defects remains the main motivation for review, reviews are less about defects than expected and instead provide additional benefits such as knowledge transfer, increased team awareness, and creation of alternative solutions to problems. Moreover, we find that code and change when to use code review and how it should fit into their development process. Researchers can focus their attention on practitioners’ challenges to make code review more effective.

We present an in-depth study of practices in teams that use modern code review, revealing what practitioners think, do, and achieve when it comes to modern code review.

Since Microsoft is made up of many different teams working on very diverse products, it gives the opportunity to study teams performing code review in situ and understand their expectations, the benefits they derive from code review, the needs they have, and the problems they face.
Code Review Goals

- **Finding defects**
  - both low-level and high-level issues (requirements/design/code)

- **Code improvement**
  - readability, formatting, commenting, consistency, dead code removal, naming, coding standards

- **Identifying alternative solutions**

- **Knowledge transfer**
  - learn about API usage, available libraries, best practices, team conventions, system design, "tricks", "developer education", especially for junior developers
Code Review Goals (cont'd)

- Team awareness and transparency
  - let others "double check" changes
  - announce changes to specific developers or entire team ("FYI")
- Shared code ownership
  - openness toward critique and changes
  - makes developers "less protective" of their code
Outcomes
(200 Microsoft reviews, 570 comments)

- Most frequent: **code improvements** (29%)
  - 58 better coding practices
  - 55 removing unused/dead code
  - 52 improving readability
- Moderate: **defect finding** (14%)
  - 65 logical issues (“uncomplicated logical errors, eg., corner cases, common configuration values, operator precedence)
  - 6 high-level issues
  - 5 security issues
  - 3 wrong exception handling
- Rare: **knowledge transfer**
  - 12 pointers to internal/external documentation etc
Outcomes

- Code Improvements
- Understanding
- Social Communication
- Defects
- External Impact
- Testing
- Review Tool
- Knowledge Transfer
- Misc
Aside: Philosophy

- One definition of the source of unhappiness is unrealized desires
  - You are unhappy when you desire reality or your experiences to have property X but they do not
  - Buddhism: “craving is the cause of all suffering”
- You can either change what you want
- Or try to change reality / your experiences
- Both are usually very difficult!
Expectation/Outcome Mismatch

- Low quality of code reviews
  - Reviewers look for easy errors (formatting issues)
  - Miss serious errors
- **Understanding** is the main challenge
  - Understanding the reason for a change
  - Understanding the code and its context
  - Feedback channels to ask questions often needed
- No quality assurance on the outcome
Formal Code Inspections

- In a formal code inspection a group of developers meets to review code or other artifacts
  - Popularized by IBM in the 1970s, broadly adopted in the 1980s, subject of much research
- Viewed as the most effective approach to finding bugs
  - 60-90% of bugs were found with inspections
- Very expensive and labor-intensive
Inspection Team and Roles

• Typically 4-5 people (at least 3 if “formal”)
  - Author
  - Inspector(s)
    - Find faults and broader issues
  - Reader
    - Presents the code or document at inspection meeting
  - Scribe
    - Records results
  - Moderator
    - Manages process, facilitates, reports
Inspection Process

Planning → Overview → Preparation → Meeting → Rework → Followup

Moderator

Inspectors (one scribe, one reader, one verifier)

Author
Inspection Steps

- Planning (select Moderator)
- Overview (brief) - Author presents context in meeting
- **Preparation** (1-2h) - Every reviewer inspects the code separately
- **Meeting** (1h)
  - Reader presents the code
  - All reviewers identify issues
  - Meetings only discover issues, do not discuss solution or whether it really is an issue
- Rework
- Followup (Verifier checks changes)
Inspection Checklists

- Reminder of what to look for
- Include issues detected in the past
- Preferably focus on few important items
- Examples:
  - Are all variables initialized before use? Are all variables used?
  - Is the condition of each if/while statement correct?
  - Does each loop terminate?
  - Do function parameters have the right types and appear in the right order?
  - Are linked lists efficiently traversed?
  - Is dynamically allocated memory released?
Process Details

- **Authors do not explain or defend** the code - not objective
  - Author != moderator, != scribe, != reader
  - Author observes questions and misunderstandings and clarifies issues if necessary
- **Reader (optional) walks through the code line by line, explaining it**
  - Reading the code aloud requires deeper understanding
  - Verbalizes interpretations, thus observing differences in interpretation
Social Issues: Egos in Inspections

• Authors *should* separate self-worth from code

• Identify defects, not alternatives; do not criticize authors
  • “you didn’t initialize variable x” → “I don’t see where variable x is initialized”

• Avoid defending code. Avoid discussions of solutions or alternatives

• Reviewers should not “show off” as smarter

• Author decides how to resolve defects
Social Issues: Inspection Incentives

• Meetings should not include management
• Do not use code reviews for HR evaluations!
  • “finding more than 5 bugs during inspection counts against the author”
  • Leads to avoidance, fragmented submission, not pointing out defects, holding pre-reviews
• Responsibility for quality with authors, not reviewers
  • “why fix this, reviewers will find it”
• cf. lecture on Metrics and Incentives
Root Cause Analysis

- An overarching goal is look beyond the immediate puzzle
- Identify way to improve the development process to avoid this problem in the future
  - Restructure the development process
  - Introduce new policies
  - Use new development tools, languages, analyses, etc.
- cf. “definition of insanity”
When to Inspect

- Inspect before **milestones**
- Incremental inspections during development
  - * Earlier* often better than later: smaller fragments, chance to influence further development
  - Large code bases can be **expensive** and frustrating to review
    - Break down, divide and conquer
    - Focus on critical components
    - Identify defect density in first sessions to guide further need of inspections
Guidelines for Inspections

• Collected over many companies in many projects and experiments

• Several metrics are easily measurable
  • Effort, issues found, lines of code inspected, etc.

Focus Fatigue

Recommendation: Do not exceed 60 minute session
Inspection Speed

Defect Density vs. Inspection Rate

Above 400 LOC/h reviews get shallow
Recommendation: Schedule fewer than 400 LOC for a 1h review session
Most issues found during preparation, not in meeting
Suggested synergy seems to have only low impact
Claim: Defects found in meetings often more subtle
Self-Checks Matter

Effect of Author Preparation on Defect Density

Authors have self-checked documents before inspection.
Inspection Accuracy

• About 25% of found issues are **false positives**
  • We'll return to this issue later in the course: it turns out humans are not perfect ...

• Avoid discussing during meeting

• Confusion during meeting is an indicator that document could be clearer
  • For maintainability, if someone says “I don't think the code does X”, it does not actually matter if the code does X or not!
Inspections vs. Reviews: Costs

• Formal inspections and modern code reviews
  • Formal inspections very expensive (about one developer-day per session)
  • Passaround distributed, asynchronous
• Code reviews vs testing
  • Code reviews claimed more cost effective
• Code reviews vs not finding the bug
Code Review by Formality

- Ad hoc review
- Passaroud ("modern code reviews")
- Pair programming
- Walkthrough
- Inspection

(When should you use which type?)
## Review Type and Differences

<table>
<thead>
<tr>
<th>Review Type</th>
<th>Planning</th>
<th>Preparation</th>
<th>Meeting</th>
<th>Correction</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal Inspection</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Walkthrough</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Pair Programming</td>
<td>Yes</td>
<td>No</td>
<td>Continuous</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Passaround (modern code review)</td>
<td>No</td>
<td>Yes</td>
<td>Rarely</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Ad Hoc Review</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Studies, Claims, Results

- **Raytheon** review study
  - Reduced “rework” from 41% of costs to 20%
  - Reduced integration effort by 80%
- Paulk et al. - costs to fix a *space shuttle* software
  - $1 if found in inspection
  - $13 during system test
  - $92 after delivery
- **IBM** - 1h of inspection saves 20h of testing
- R. Grady - efficiency data from **HP**
  - System use 0.21 defects/h
  - Black box testing 0.28 defects/h
  - White box testing 0.32 defects/h
  - Reading/inspection 1.06 defects/h
Questions?

- Homework 1b, 1c, 1d all due!