

Software Analytics in Practice: Approaches and Experiences

Dongmei Zhang
Senior Researcher/Research Manager
Software Analytics group, Microsoft Research Asia

December 16, 2012

Outline

- Software Analytics
- Experience sharing in practicing Software Analytics
- Example projects

New Era ...New Opportunities...

Software itself is changing ...

The way people use software is changing ...

How software is built and operated is changing...

Scope of software development & tools has naturally expanded...

Software Analytics Group @ MSRA

Utilize data-driven approach to help create highly performing, user friendly, and efficiently developed and operated software and services.



Research Topics

Vertical



Horizontal

Information Visualization

Analysis Algorithms

Large-scale Computing

Technology Pillars

Software Analytics

Software analytics is to enable software ***practitioners*** to perform data exploration and analysis in order to obtain ***insightful*** and ***actionable*** information for data-driven tasks around software and services.

Software Analytics in Practice

- Getting real
 - Working on real data
 - Addressing real problems
 - Building real tools
 - Making real impact
- Experience sharing
 - Engagement of practitioners
 - Walking the last mile
 - Combination of expertise

Engagement of Practitioners

- Broad range of practitioners
 - Developers, testers, program managers, UI designers, customer support, operators...
- Solving their problems
- Champions in product teams
- Timing
- Culture

Walking the Last Mile

- Targeting at real scenarios
- Trying out tools has cost
- “It works” is not enough
 - Performance
 - Usability
 - Customizability
 - Predictability
- Feedback & improve -> iterative process
- Getting engineering support

Combination of Expertise

- Research capabilities
- Engineering skills to build systems
- Visualization & design lead to ease of use
- Project management
- Communication

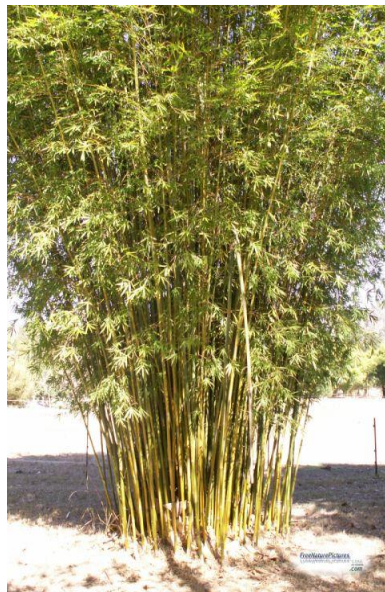
Example Projects

- Code Clone Analysis
 - Yingnong Dang, Song Ge, Gong Cheng, Weipeng Liu, Dongmei Zhang
- StackMine
 - Shi Han, Yingnong Dang, Song Ge, Dongmei Zhang

XIAO – Code Clone Analysis

四十年来画竹枝，
日间挥写夜间思；
繁冗**削**尽留清瘦，
画到生时是熟时。

– [清] 郑板桥



“削”
→
XIAO



- 削 (XIAO) means “trimming” in Chinese
- Similar spirit between bamboo painting and programming

XIAO: Code Clone Analysis

- Motivation
 - Copy-and-paste is a common developer behavior
 - A real tool widely adopted at Microsoft
- XIAO enables code clone analysis in the following way
 - High tunability
 - High scalability
 - High compatibility
 - High explorability



Code Clone Detection Experience at Microsoft

Yingnong Dang, Song Ge, Ray Huang and Dongmei Zhang
Microsoft Research Asia

yidang.songge@rayhuang.dongmeizhang@microsoft.com

ABSTRACT

Cloning source code is a common practice in the software development process. In general, the number of code clones increases in proportion to the growth of the code base. It is challenging to proactively keep clones consistent and remove unnecessary clones during the entire software development process of large-scale commercial software. In this position paper, we briefly share some typical usage scenarios of code clone detection that we collected from Microsoft engineers. We also discuss our experience on building XIAO, a code clone detection tool, and the feedback we have received from Microsoft engineers on using XIAO in real development settings.

Fix Bugs: Once a bug is identified in a piece of code with duplicated copies, it is desirable to have the ability to fix all of them at once. This scenario is beneficial to multiple stages of the development process as long as there are bug fixing tasks; for example, during the feature implementation stage, stabilization stage and post-release maintenance stage.

Footprint Reduction: Code clones can be found at various degrees for different product teams we have worked with in Microsoft. Some teams are keen on reducing the memory footprint of their components; they look for every possible opportunity to achieve this goal. Removing code clones is one of the important actions they want to take.

[IWSC'11 Dang et.al.]

Comprehensive Solution

Quality gates at milestones

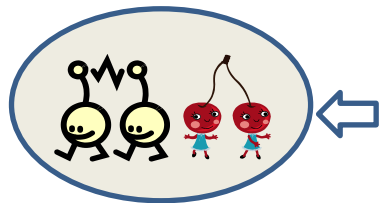
- Architecture refactoring
- Code clone clean up
- Bug fixing



Online code
clone search

Post-release maintenance

- Security bug investigation
- Bug investigation for sustained engineering



Offline code
clone analysis

Development and testing

- Similar issue check before check-in
- Reference info for code reviewer
- Supporting tool for bug triage



Adoption in Microsoft

- More than 900 downloads
- Gaining overall understanding of copy-and-paste clones in a codebase
- Finding potential bugs & refactoring opportunities
- Adding custom parsers

More Secure Microsoft Products



Code Clone Search service integrated into workflow of Microsoft Security Response Center



Over 400 million lines of code indexed across multiple products

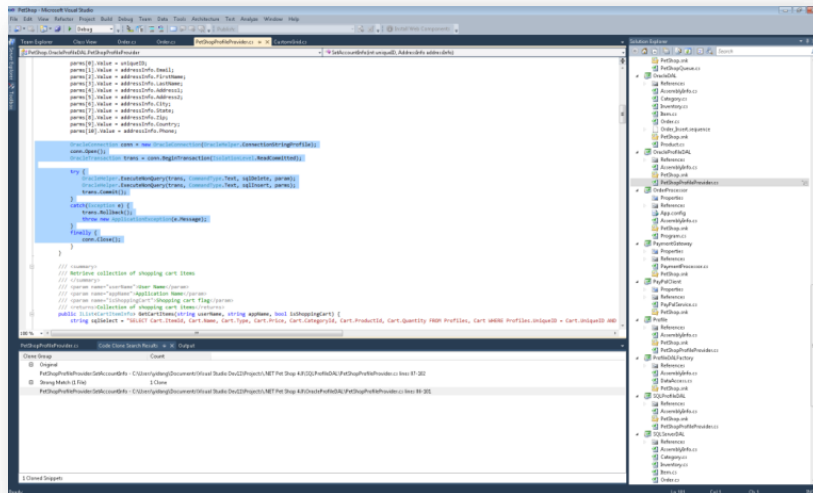


Real security issues proactively identified and addressed

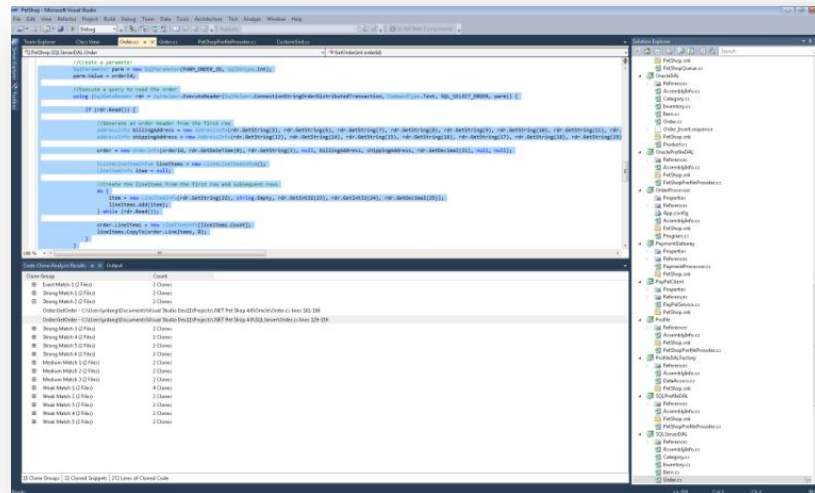
Benefiting Developer Community



Available in Visual Studio vNext

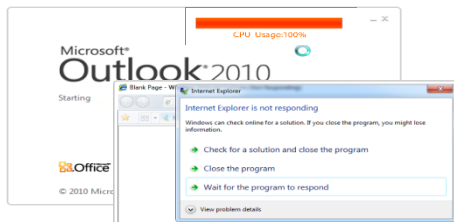


*Searching similar snippets
for fixing bug once*



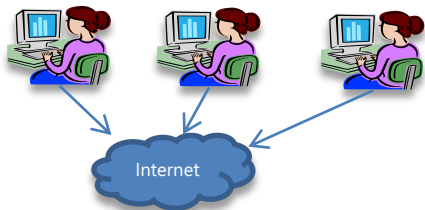
*Finding refactoring
opportunity*

StackMine: Towards Flawless OS Performance



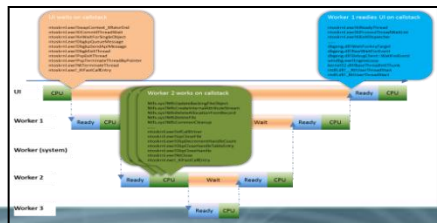
OS performance in the real world

- One of top user complaints
- Impacting large number of users every day
- High impact on usability and productivity



Challenges

- Large-scale trace data
- Highly complex performance analysis at OS level
- Combination of machine learning and domain knowledge



Problems

- Unknown issue discovery
- Issue prioritization
- Scalable to large number of traces

Technical Highlights

- Machine learning for system domain
 - Formulate the discovery of problematic execution patterns as callstack mining & clustering
 - Systematic mechanism to incorporate domain knowledge
- Interactive performance analysis system
 - Parallel mining infrastructure based on HPC + MPI
 - Visualization aided interactive exploration

Impact



"We believe that the MSRA tool is highly valuable and much more efficient for mass trace (100+ traces) analysis. For 1000 traces, we believe the tool saves us 4-6 weeks of time to create new signatures, which is quite a significant productivity boost."

- from Development Manager in Windows

Highly effective new issue discovery on Windows mini-hang



Continuous impact on future Windows versions

Suggested Actions

- Get research problems from real practice
- Get feedback from real practice
- Collaborate across disciplines
- Collaborate with industry

Summary

- Scope of software development & tools has naturally expanded
- Software Analytics – Insightful & actionable
- Experience sharing
 - Engagement of practitioners
 - Walking the last mile
 - Combination of expertise

Q & A

<http://research.microsoft.com/groups/sa/>