MSR 2012 @ ICSE

Table of Contents

- The evolution of software
- Do Faster Releases Improve Software Quality?
- Security vs Performance Bugs in Firefox
- Some of my thoughts
  - Separation of commits based on Semantic analysis
  - About the slide systems used today
  - Microsoft is MEANING

Mining Software Repository 2012 @ ICSE

I participated MSR of this year. We came to University of Zurich early in the morning. The registration got something wrong when it seems that Swisses cannot tell the difference among Asians so that name cards of 3 Chinese with family name of Yang are misplaced. And also the organization field of Hotta was "Japan, Japan", as if he represented the Japan.

MSR(MicroSoft Research) talk @ MSR(Mining Software Repositories)

The first talk was the keynote given by Mrs Zhang from MSR(MicroSoft Research @ Asia), so it turned out to be MSR gave keynote of MSR. The talk was about Software Analysis and their clone detection tool called XIAO. XIAO was a clone detector developed by MSRA which can be used as a plugin for Microsoft Visual Studio. XIAO has two part, or system state: the statics state analysis all the clones which didn't consider the running time, while the dynamic state need real time response. The thing I need to develop for Samsung is something like dynamic mode. I wanted to know more about the internal details about XIAO but the talk was finished there.

Towards Improving BTS with Game Mechanisms

The contents of this talk is very much like this blog:

http://www.joelonsoftware.com/items/2008/09/15.html

The talk discussed whether the same game mechanism can be applied to the things like issue tracking or similar. From my point of view, it is useless to use game mechanism in this situation. The reason that stackoverflow can success lies on that they just captured the use of fade system in opensource community, as all hackers like to be approved as great hacker, as what is happening in Wikipedia. Whether the same theory can be applied in issue tracking systems inside a internal company is questionable. Although MSDN has basic the same structure as Wikipedia, the content of MSDN and Wikipedia have different involvement of users. So I myself didn't approve this research.

GHTorrent

They slide of this talk can be found from here: http://www.slideshare.net/gousiosg/ghtorrent-githubs-data-from-a-firehose-13184524

Data exporter for github. Main part of data of Github, namely the hosted code, are already exposed as git repos, and wiki of repos are stored in

git repo. So the aim of this project is to expose other data such as issues, code comments, etc. The project access github api and fetch the needed data as distributed system in order to overcome the limitations of the github api. The project will provide download history as torrents. The json data from github api is stored as bson in MongoDB and the parsed data is stored in MySQL with schema.

From my point of view, it will be better if the format of data can be uniformed and all data are stored in the git repo as wiki pages. As the history stored in git repo is more nature, and using

```
git blame
```

to trace author of code comments should also be more useful. Of course it is harder to read and write the raw data of git as we need more understanding of the internal format of git. Maybe only people from github can do this.

Topic Mining

I can not understand the two parameters, DE, AIC, used in this research, study this later. The experiment target of this research are Firefox, Mylyn and Eclipse. They are trying to analysis the identifiers and comments from source codes in software repos and find the relationship between topics and bugs, like what kind of topics are more likely to contain buggy codes.

The result of this research is not so clear. Such as it said that the core functions of Firefox have more bug reports, but it said no reason about this. Maybe this only means that the core features are well tested, rather than that the core features are more buggy.

But the slides showed by author are pretty and easy to understand.

The evolution of software

The keynote talk of the second day. It is about how should we combine the social media with software development. Maybe this is the reason why Github succeeded. In the talk she told about

accessing tags, uBlogs, blogs etc. directly from Integrated Development Environments, or should we need cloud IDE such as Cloud9.

Do Faster Releases Improve Software Quality?

Used Firefox as example.

The conclusion is that faster releases will lead to more bugs and more frequent crash, but bugs are get fixed more quickly and user will switch to new released more quickly.

Security vs Performance Bugs in Firefox

Performance bugs are regression, blocks release.

Some of my thoughts

Separation of commits based on Semantic analysis

The user of some tools (such as git) are not following the design purposes of these tools which brings some difficulty to MSR. For example git has a prefect branch system, so it is desired for users of git to commit per topic. Commit per topic means that user send a commit for a single implementation of a feature or a bug fix, etc. If it is difficult to contain all modifications in a commit, then it should be in a separate branch and merged into master branch. But actually user tends to send very large commits, that contains many logical features, and they can not predict to open a new branch until a few commits.

Maybe this is not the fault of the user of tools, this is the tools that are not smart enough. We should separate the commits according to the semantic topics inside a commit.

About the slide systems used today

The study with title

```
Incorporating Version Histories in Infor
mation
Retrieval Based Bug Localization
```

used the slides made by beamer. It contains many equations, used many overlays are iterations, with few figures, is a typical beamer slide. It also used mindmap very well.

There are at least 3 slides that are made by beamer today.

The study with title

```
Towards Improving Bug Tracking Systems w
ith
Game Mechanisms
```

presented with prezi. It have many pictures and many transitions. But because of it is made by prezi, there are no headers and footers so no page numbers and section titles etc. This is not so convenient in such a official occasions because people need to refer to the page number in question session.

There are at lease 6 presents used Apple Keynote. It is really difficult to tell the difference between slides made by PowerPoint and Keynote. 2 of them used the default theme of keynote.

The rest are using PowerPoint. Mrs Zhang from Microsoft used PowerPoint but her slides looks like beamer very much such as the usage of footer and header and overlays. If these are made by PowerPoint that will involve many manually operations.

It is worth to mention that the slides of a study with title

```
Green
Mining: A Methodology of Relating Softwa
re Change to Power Consumption
```

are all

```
badly
```

drawn hand paintings. The effect of these slide are well received, they are green and clean and cute. You can refer to the following animation for the effect but it is not exactly the same version with what we saw :

http://softwareprocess.es/a/greenmining-presentatation-at-queens-20120522.ogv

Microsoft is MEANING

It is not a news. But Microsoft is the sponsor of Mining Challenge, and the prize of this challenge will be Xbox and Kinect and the topic of this year is:

```
Mining Android Bug
```

I see what you are doing there Microsoft ……