

# MSR 2012 @ ICSE

## 目次

- Mining Software Repository 2012 @ ICSE
  - MSR(MicroSoft Research) talk @ MSR(Mining Software Repositories)
  - Towards Improving BTS with Game Mechanisms
  - GHTorrent
  - The evolution of software
  - Do Faster Releases Improve Software Quality?
  - Security vs Performance Bugs in Firefox
  - 思いつき
    - topic に基づいて commit の分析と分割
    - 今回に皆使っている slide のシステム
    - マイクロソフトは腹黒っ子!

# Mining Software Repository 2012 @ ICSE

今年の MSR を参加しました、会場はチューリッヒ大学にあります。朝早く大学に着いて、登録するときちょっと事情がありました。スイス人は明らかに中国人の名前をわからないから、3つの中国からの楊(Yang)の名札を間違えた。そして堀田先輩の名札に "Japan, Japan" になって、日本代表になった。

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

まず一番目の Keynote はマイクロソフトアジア研究院 (MicroSoft Research @ Asia ,MSR Asia) の Zhang 氏が発表する、こうして MSR が MSR に発表するになった。

Zhang の発表は Software Analysis と XIAO の2つの紹介です。XIAO はマイクロソフトが開発した Code Clone Detector、ある会社が私達に任せるのもこのようなシステムです。もっと詳しく知りたいが、実装に関わるものは言ってなかった。

## Towards Improving BTS with Game Mechanisms

この内容は基本的にこのブログに書いています：

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

同じ理論を Issue Tracking とかに応用できるかを言いました。個人的にこれは 意味ない気がします。stackoverflow の成功は Open Software Community にもと もとある名誉システムを具現化したです

から、それを会社の中に応用するのは 難しい気がする。

## GHTorrent

この研究のスライドはこちら

に: <http://www.slideshare.net/gousiosg/ghtorrent-githubs-data-from-a-firehose-13184524>

Data exporter for github. Github の主なデータはコード、それは既に git から アクセスできます、wiki は git として保存しているからそれも含まれている。ですからこのプロジェクトの目的は他のデータを表せる、つまり issues, commit comments など。このプロジェクトは github api を通じて、分布システムとして api の制限を超える、そして torrent の形で歴史を download できます。元のデータ は bson として MongoDB の保存して、Schema を追加したデータは MySQL に保存する。

わたしの意見では、データを git の repo の形で保存するの方がいいかもしれない。今の wiki のように、そして gitolite も全てのデータを git 自身の中に保存している。

## The evolution of software

二日目の keynotes, social media をソフトウェア開発に巻き込めるについて 話しました。もしかしてこれは Github の成功の理論かもしれない。IDE の中に social media のアクセスを欲しいと言いました。

## Do Faster Releases Imporve Software

# Quality?

Firefox を例として研究しました。

結論としては、早い発行は bug を多く持たされ、crash がもっと頻繁になるが、bug の修復も早くなって、そしてユーザー側はもっと早く新しい発行に移動する ことをわかりました。

## Security vs Performance Bugs in Firefox

性能に関する bug は regression テストが要る、そして発行を阻止する。

---

## 思いつき

### topic に基づいて commit の分析と分割

よく使うツール（例えば git）のユーザーはツールの設計者の意図に従って ツールを使うことはない、設計者が思った用途以外にも使っていることが多い、それは Mining に対しては色々困難を持たされています。例えば git には完璧な branch 機能がある、通常に git のユーザーが一つの commit に一つの機能を実現 してほしい、例としては bug の修復とか、機能の追加とか。それは難しいなら branch を使って、一連の commit を一つの branch になって、一つの branch に一つの機能を実現してほしい。それなのに、現状では、沢山の編集を一つの commit に含まれていて、後の管理とか情報の収集とかが困難になってしまう。

それはユーザーの悪いと思わない、ツールの方がもっと頑張らない

とユーザー は正しく使えない。もし commit の時、自動的に commit の内容を分析して、その中に topic によって分けて、ユーザーに推薦するのをてきたらいいなあ、と思っています。このように一つの commit を多くに分割したら、commit の履歴 をもっと見やすくなって、続いて分析とかも便利になるはずです。

## 今回に皆使っている slide のシステム

タイトルは

Incorporating Version Histories in Information Retrieval  
Based Bug Localization

の人が使っているのは beamer です。数式が多くて、overlay も多くて、iterate も多い、図だけ少ない、典型的に beamer に作れそうな スライドです。mindmap の使いもうまい。今日の一日に少なくとも 3 個の slide は beamer で作られています。

タイトルは

Towards Improving Bug Tracking Systems with Game Mechanisms

の人は prez を使いまして、図が多くて、transition も多い。但しスライドとしては必要なページ数とかがなくて、このような国際会議の場合にはもっと 工夫をした方がいいかもしれな。

少なくとも六人以上は Apple の Keynote をつかていもう。Keynote による作った スライドは Powerpoint のになかなか区別しがたいですが、その中に二人は default の theme を使ったからわかります、他の人は PPT に決してありえない アニメーションを使っていますから、多分 keynote。

残りは勿論 Powerpoint です。MSRA の張さんが作ったのは powerpoint なんですけど、すごく beamer の感じがします、例えば header と footer の使い方とか、overlay 見たいのものでページの item を一つずつ展開するとか。それらを全部 powerpoint で作るのは相当手間がかかりそうです。

ちなみに言いたいのは一つタイトルは

Green Mining: A Methodology of  
Relating Software Change to Power Consumption

のスライドは全部

下手

な手描きの漫画で表せている、火狐のアイコンさえ手描きする、効果は意外に評判がいい。省エネでグリーンで環境にいいで可愛らしい。具体的な効果は下のリンクから見えます、現場で見たのは別のバージョンなんですけど：

<http://softwareprocess.es/a/greenmining-presentation-at-queens-20120522.ogv>

## マイクロソフトは腹黒っ子!

まあ大したニュースではないですけど、MSR2012 の Mining Challenge のスポンサー はマイクロソフトで、商品は Xbox と Kinect ですけど、今年のチャレンジのテーマは：

Mining Android Bug

マイクロソフトの殺意を感じました。

