Process Mining Software Repositories with FRASR

Bram Schoenmakers

June 15, 2011

Capita Selecta (2IS99)
Supervised by: A. Serebrenik

Introduction

- FRASR: Framework For Analysing Software Repositories
- Graduation project of Wouter Poncin (2010)
- Contribution to CSMR 2011
- Capita Selecta: extend and improve FRASR
Purpose

▶ Try to answer software engineering related questions such as:
  • Who are the architects in a project and who is doing the real work (implementation)?
  • Is there a relation between the day of the week and the number of bugs introduced?
  • Has the V-model been applied during a software project?
▶ Apply process mining techniques to data in software repositories.

FRASR input

▶ Version control systems
▶ Bug trackers
▶ Email messages
▶ Forum posts
▶ Trac Wiki pages
Developer matching

Schematic output of FRASR

Steve B.
- Check-in revision: 888
- Email: [RFC] Bugfix in iPushButton.cpp

Steve J.
- Check-in revision: 890
- Email: Re: [RFC] Bugfix in iPushButton.cpp
- Bug closed: 1338

Linus T.
- Check-in revision: 891
FRASR output

- CSV
- MXML

Git integration

- Used by small and large software projects
- Ability to select which branches to analyse
- Git tracks more information than CVS / Subversion:

  commit 30b4caf5d73af5c99cf1b2b46496d8bc35330992
  Author: Li Zefan <lizf@cn.fujitsu.com>
  AuthorDate: Wed Jun 8 03:56:44 2011 +0000
  Commit: Chris Mason <chris.mason@oracle.com>
  CommitDate: Sat Jun 11 08:31:55 2011 -0400

  Btrfs: use join_transaction in btrfs_evict_inode()
  [...] 

  Reported-by: Tsutomu Itoh <t-itoh@jp.fujitsu.com>
  Signed-off-by: Li Zefan <lizf@cn.fujitsu.com>
  Reviewed-by: Josef Bacik <josef@redhat.com>
  Signed-off-by: Chris Mason <chris.mason@oracle.com>
Content

- Make FRASR aware of file contents
- Extract metrics, ignore uninteresting commits, ...
- Interface to external tools written

Current status

- Content almost integrated
- Goal for this capita selecta: cope with the Linux kernel source
Future work

- More data sources
- Write interfaces to other tools
- Optimization
- Detach GUI
- GUI improvements w.r.t. usability
- ...

Questions

Questions?