Mining Open Source Software Repositories

Peter C Rigby

Introductions

- Professor: Peter Rigby
- Professor in Software Engineering
- Postdoc work at McGill
- PhD at the University of Victoria
- Bachelor in Software Engineering at University of Ottawa

My Research Areas

- Empirical Software Engineering
- Software Analytics and Mining
- Statistical Machine Translation of Code
- http://users.encs.concordia.ca/~pcr

Possible research projects

- Effect on number of defects
  - Continuous integration
  - Turnover
  - Refactorings
  - Social interactions among devs
  - Complexity of code

Mining Data

- Without empirical grounding, we just have opinions
- Mining large data can be applied to types of data sets
- ISO and other standards require tracking progress

What is Open Source Software?

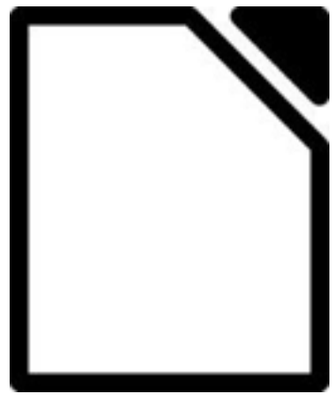
Products

Proprietary

iPhone

Microsoft® Office

Open Source



OSS Products

- Do they threaten software development companies and paid developers?
- Few people know how to make money from OSS (Business models)
- Few people know how to integrate legally use OSS in proprietary development (licenses)

Introductions

- Name
- What interests you about the course
  - Data Mining or Open Source or both?
- What would you like to study?

Lecture Structure

- Lecture on OSS and development practices
- Presentation/Discussion of data mining and research papers

Grades

- Software Practices Exam 40%
- Bug fix on an OSS Project 20%
- In class discussions and simple summaries of papers 15%
- Research paper including model of data 25%

Grade Alternative 1

- For Researchers
- ~~Bug fix on an OSS Project 20%~~
- Research paper including model of data 45%
- Software Practices Exam 40%
- In class discussions and simple summaries of papers 15%

Grade Alternative 2

- For Developers and Practitioners
- Bug fix on an OSS Project 45%
- ~~Research paper including model of data 45%~~
- Software Practices Exam 40%
- In class discussions and simple summaries of papers 15%

Research Project

- Choose a problem (and related literature)
- Choose data and software product to study
- Analyze data (weka or R)
- Write a four page, double column report

Presentation

- Everyone will pick and present their favorite paper (related to research project)
- I will present at least one research paper per week, and we will discuss it

Bug Fix Project

- Pick a small OSS project to work on
- Fix a bug and send it the the maintainers
- Contrast the development process of this project, with the ones you are mining for the research paper

Software Practices

- ISO vs Agile vs OSS
- Practices
  - Integration, testing, review, planning, etc
- Tools
  - GitHub, ReviewBoard, etc