Methodology of Multi-Criteria Comparison and Typology of Open Source Projects

Fedir RYKHTIK, October 22, 2018, Edinburgh, UK
Sponsors

- **Diamond**: Intel, Microsoft
- **Platinum**: IBM, SUSE
- **Gold**: AWS, Google, Nutanix, Cloud Native Computing Foundation, Oracle, Red Hat, Sumo Logic, VMware
Fedir RYKHTIK

• Building open source web since 2007
  – Back-end developer
  – Independent researcher
  – DevOps / SA
• CTO @AgenceStratis since 2015
OSS today
OSs today

Github only

• 96 M+ repositories
• 40% more than last year
• 31 M+ developers
Problem of choice & following
Multi-Criteria Comparison
What is a good OSS project?

Reliable
Useful
Maintained
Secure
Different layers have own metrics

- Core team
- Ecosystem maintainers
- Project integrators
- Clients
Software groups of metrics (A-Z)
Metrics > Agile

• Lead time (period of time between ticket creation and resolving)
• Open / closed rate
Metrics > Documentation

• Technical documentation coverage
• Articles & manuals
• Books
Metrics > Marketing

- Social networks marketing
- Search engine optimization
Metrics > Performance

- Volume of servers/CPU/... required
- Execution speed
- Supported charge
Metrics > Production

- Active days
- Tasks scope
- Code churn
- Apps crash rate
Metrics > QA

- Number of bugs
- Frequency of bugs
- Returning bugs
Metrics > Security

- Time / ressources to find a security bug
- Time / ressources to fix it
Metrics > Size-oriented

- Number of code lines
- Number of bugs per 1000 code lines
- Number of classes and interfaces
- Number of commits
Metrics > Usage

- Accessibility
- Number of features
- Simplicity of usage
- Unique features
OSS specific metrics
Metrics > OSS > Author

- Notoriety / Experience
- Involvement
Metrics > OSS > Community of contributors / integrators

- Social ranking (stars)
- Downloads
Metrics > OSS > Community of developers

- Community size
- Active forkers
- Notoriety / Experience
- Returning contributors
- Medium contribution period
Metrics > OSS > Languages

- Number of used languages
- Popularity of the language
Metrics > OSS > Rhythm

- Last commits
- Regular maintenance
- New versions
How do we collect it?
Data collection

- Code analysis (LoC, coding conventions, coupling, deps)
- Unit testing (code coverage, number of scenarios)
- Versioning systems (code, contributors, branches, tags)
- Using application (confirming functional perimeter)
- Social networks (community, feedback)
- Search engines (buzz, books, materials)
- Bugtracker statistics (bugs, maintainers activity)
- Benchmarking (load, endurance, stress, limits)
- Pentesting (automated, manual)
Project quality index
Counting project quality index

\[ Q_{total} = \sum_{m} Q_{m} \]

\( Q_{m} \) - quality index of current metric
Counting project quality (custom)

\[ Q_{total} = \sum_{m} (Q_m \times C_m) \]

- \( Q_m \) - quality index of current metric
- \( C_m \) - optional custom coefficient for current metric
Analyzer prototype
fedir/ghstat (WIP)

- [https://github.com/fedir/ghstat](https://github.com/fedir/ghstat)
  - Statistical multi-criteria comparator for Github's projects
- What does it do
  - Collects statistics from Github
  - Calculate additional metrics
  - Gives points and ranks projects
- Statistics: Name, URL, Author, Author's location, Main language, All used languages, Number of languages, Description, Total code size, License, Author's followers, Top 10 contributors followers, Created at, Age in days, Total commits, Total additions, Total deletions, Total code changes, Last commit date, Commits/day, Average contribution period by contributor in days, Medium commit size, Total releases, Stargazers, Forks, Contributors, Active forkers(%), Returning contributors (more than 4 weeks), Open issues, Closed issues, Total issues, Issue/day, Closed issues (%)
fedir/ghstat statistics examples

https://github.com/fedir/ghstat/tree/master/stats

- Open source programming languages
- Web frameworks
- Content management systems
- ...

THE LINUX FOUNDATION OPEN SOURCE SUMMIT
fedir/ghstat - placements

- Placement by popularity
- Placement by age
- Placement by total commits
- Placement by total tags
- Placement by top 10 contributors followers
- Placement by closed issues percentage
- Placement by commits by day
- Placement by active forkers column
fedir/ghstat results sample
fedir/ghstat > future features

- offline git repository scanner
- social connectors (gitlab, bitbucket, gitea...)
- more metrics
  - lots of stuff to do
Live demo
$ ./ghstat -r Microsoft/MS-DOS,freebsd/freebsd,redox-os/redox -f os.csv
$ ./ghstat -r torvalds/linux
Summary
Results

- OSS has additional domain specific metrics
- Using multi-criteria comparison methods we can choose / control states of packages health by our needs
- Using statistical analyzers gives You more possibilities
  - You could identify interesting packages faster
  - As developer, You could identify, which project needs Your help
  - As contributor, You could identify, which project follows Your needs
Q&A
Any questions?
Related materials

• https://octoverse.github.com/
• https://en.wikipedia.org/wiki/Multiple-criteria_decision_analysis
• https://en.wikipedia.org/wiki/Analytic_hierarchy_process
• https://en.wikipedia.org/wiki/Group_decision-making
• https://en.wikipedia.org/wiki/Analytic_network_process
• https://wackowiki.org/doc/org/articles/5typesopensourceprojects
• https://techbeacon.com/top-5-software-quality-metrics-matter-right-now
• https://diceus.com/top-7-software-quality-metrics-matter/
• https://en.wikipedia.org/wiki/Programming_complexity
• https://en.wikipedia.org/wiki/Maintainability#Software_engineering
• https://github.com/fedir/ghstat
Used media resources

- https://commons.wikimedia.org/wiki/File:Big_%26_Small_Pumkins.JPG
- https://commons.wikimedia.org/wiki/File:Singapore_Road_Signs_-_Restrictive_Sign_-_Stop_-_Security_Check.svg
- https://commons.wikimedia.org/wiki/File:Green_bug.svg
- https://commons.wikimedia.org/wiki/File:Scrum_process.svg
- https://svgsilh.com/image/459225.html
- https://www.flickr.com/photos/daniel_iversen/15090961835
Thank You!

Send me feedback

https://fedir.github.io/feedback.html

@FedirFR