

Open Source Community Metrics

Tips and Techniques for Measuring Participation

LibreOffice Conference
October 2012

Dawn M. Foster
Community Manager at Puppet Labs



@geekygirl dawn

dawn@puppetlabs.com

fastwonderblog.com
puppetlabs.com

Stuff I'll Talk About

- What, why and example metrics
- Coming up with the right metrics
- Tips and techniques for collecting metrics
- Sharing metrics and highlighting community members

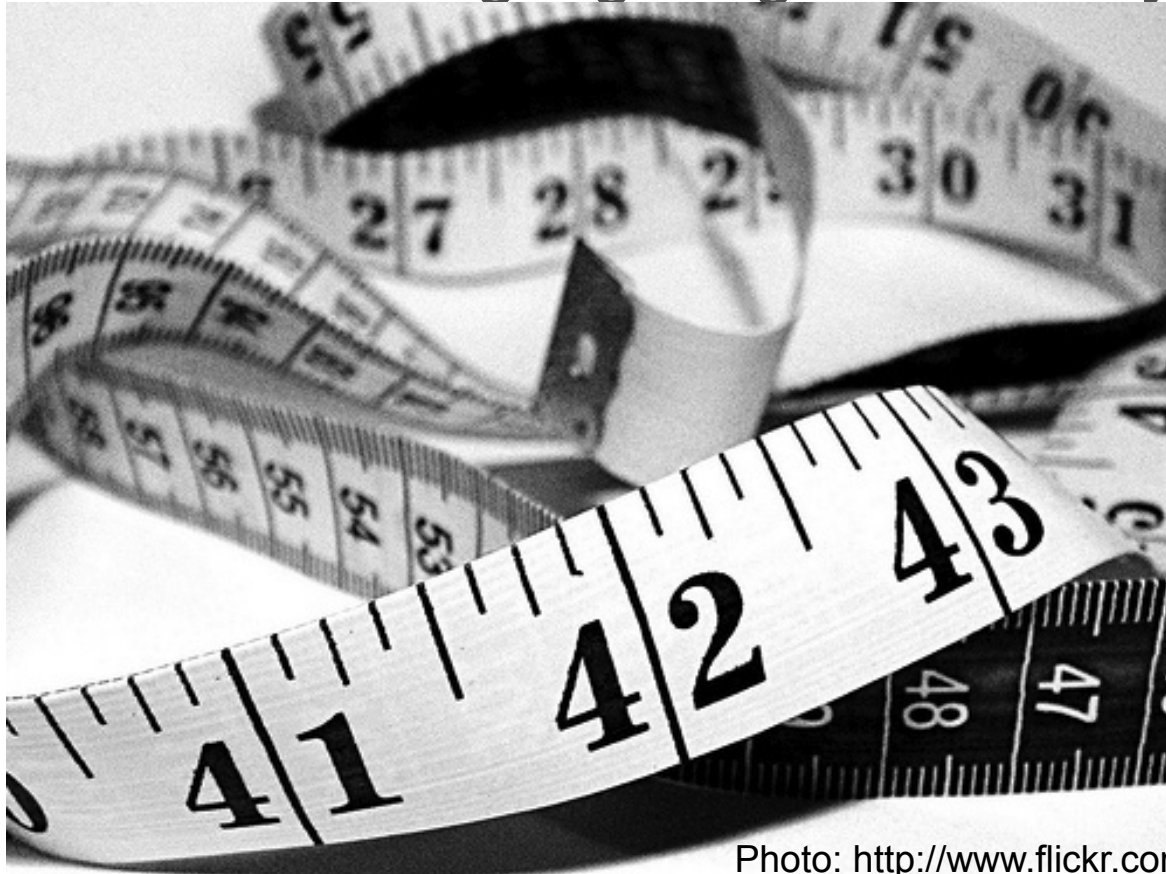


Photo: <http://www.flickr.com/photos/falcifer/3136673599>

Community Definition

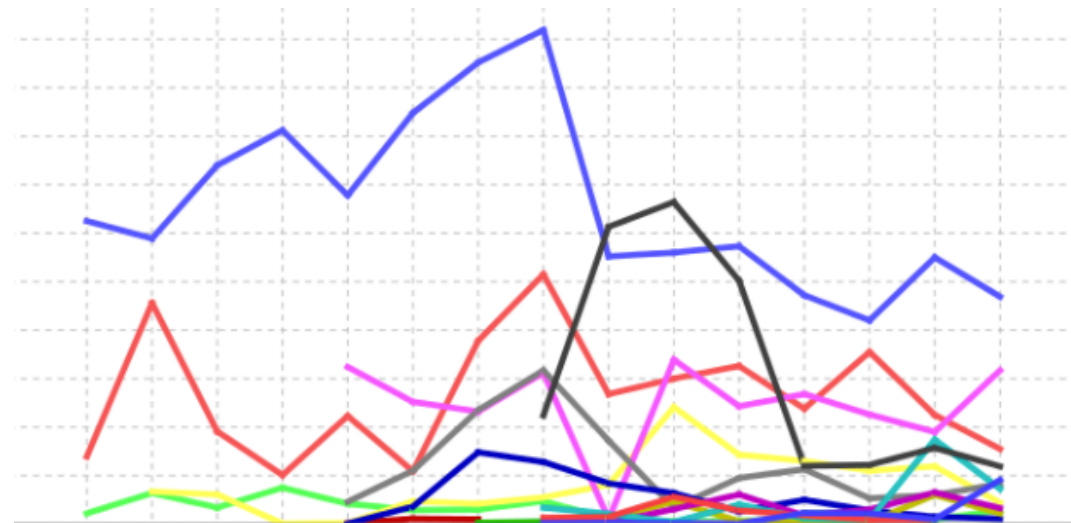
- Community includes all of the people who work on the project
 - **Product contributors:** developers, release managers, quality assurance, localization, etc.
 - **Other developers:** writing applications, modules, extensions, etc.
 - **Users:** people who run your software and provide feedback
 - **Vendors:** companies with products / services based on your project
 - **Other contributors:** promotion, moderation, documentation and more

Some people contribute as part of their employment at companies, while others contribute free time. The community includes all of the people who are working on your project.



Metrics are Useful for Open Source Projects

- Measure progress in your community over time
- Who contributes
- Where are people contributing
- Spot trends
- Gauge interest
- Learn more about key contributors
- Recognize contributions



Example: Components of an Open Source Community



Example: Puppet Metrics September 2012

Summary

4688 Members & 809 messages in Puppet-Users

877 Members and 121 Messages in Puppet-Dev

848 Nicks on #puppet IRC Channel

1625 Forge Users & 528 Forge Modules

3388 Redmine Accounts

381 Forks / 988 Watchers of Puppet

Six Month Comparison (May 2012)

3907 Members & 843 messages in Puppet-Users

751 Members and 242 Messages in Puppet-Dev

739 Nicks on #puppet IRC Channel

1012 Forge Users & 310 Forge Modules

2783 Redmine Accounts

298 Forks / 696 Watchers of Puppet

<http://PuppetLabs.com/community/metrics>

Example: June 2011 MeeGo Community Metrics Summary

- 439,963 unique people visited this month (430,070 last month).
 - 1,796,288 pageviews (1,857,907 last month); 599,573 total visits (597,658 last month)
- 25,801 people are members of MeeGo.com (was 24,019 last month)
 - Dev ML subscribers = 4772; Community = 3736; iL10N = 2607; SDK = 3197
- Mailing Lists: 5890 posts this month; 200 people posted 2+ msgs
- Forums: 2255 posts. 242 people posted 2+ messages
- New Bugs Created: 1938; Bugs Resolved: 2871
- Downloads: 52,627 1.2 Netbook, 626 1.2 IVI, 1505 1.1 N900, 3072 Tablet
- Active Users: Estimated at 800 – 1000 people.
 - Mailing Lists: 302 people with unique email addresses posted (308 last month)
 - Forums: 412 people posted at least one item (262 last month)
 - Bugzilla: 773 people performed some action (624 last month)
 - Wiki: 276 edited the wiki (205 last month)
 - IRC: 410-520 people logged into #meego simultaneously most days

<http://wiki.meego.com/Metrics>

What are the Right Metrics for YOUR Project

- Goals
 - What are your overall goals for the project?
 - How can you measure progress toward those goals?
 - What is important to you and your progress?
- Trends
 - What should you measure to recognize trends?
 - How do you recognize when something is going wrong?
 - Do you notice big improvements?



Note: I measure way too much

Mailing Lists: mlstats

Mailing List Stats is a command line tool used to analyze mailing list archives. It downloads the archives, places them in a directory and stores all the information contained in each mailing list post into a database



LibreSoftwareResearch

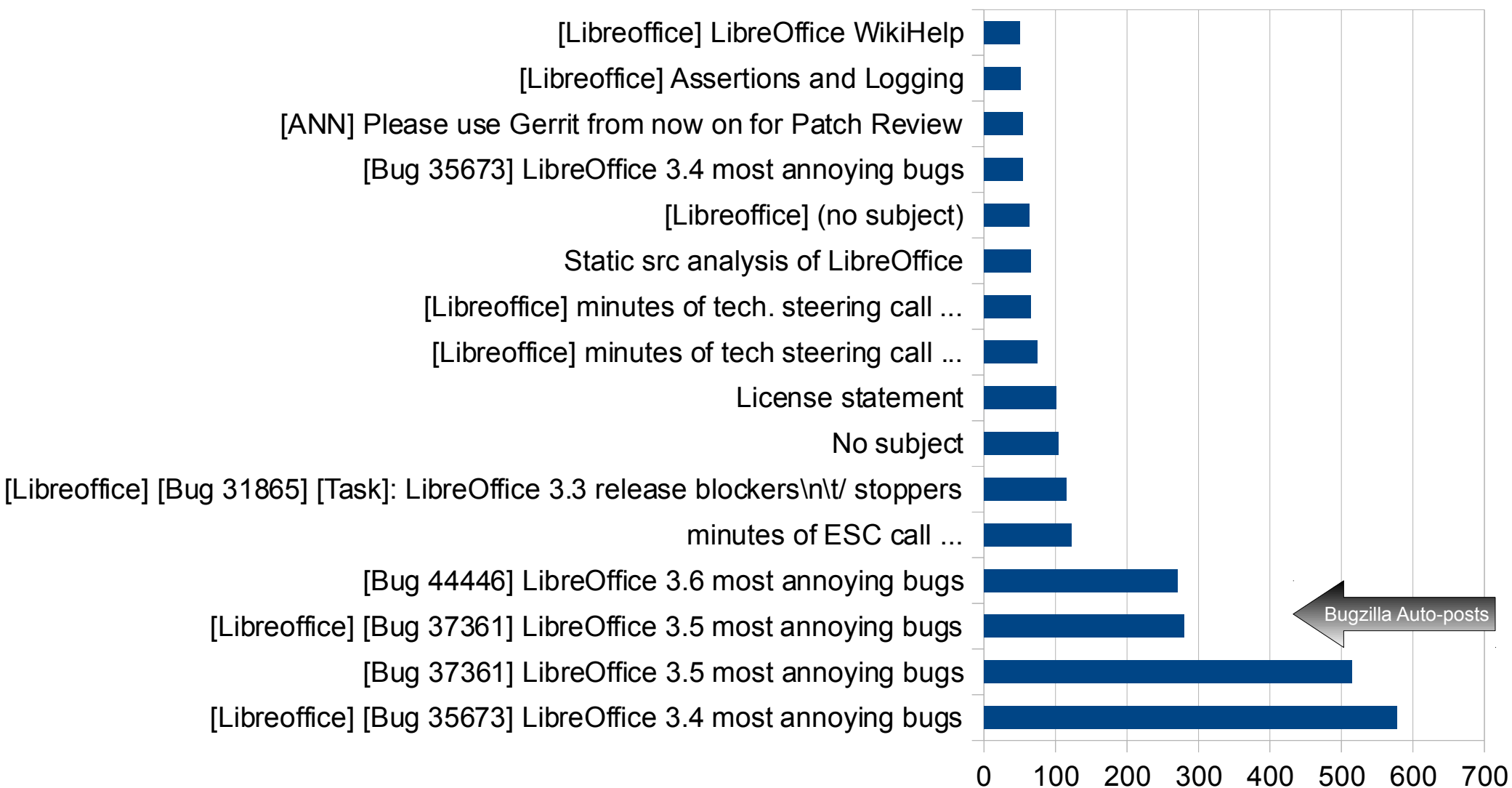
WE STUDY LIBRE SOFTWARE
ME >100% LIBRE SOFTWARE

<http://libresoft.es/tools/mlstats>

Mailing Lists: mlstats for LibreOffice Developer List

- Grab data from your mailing & store in db (repeat per ML)
 - `mlstats --db-user=root --db-password=
http://lists.freedesktop.org/archives/libreoffice/`
- Top Content Query: All or By Month
 - `select subject, count(*) as c from messages group by subject
order by c;`
 - `select subject, monthname(first_date) as m, year(first_date) as
y, count(*) as c from messages group by subject,
month(first_date) order by y, m, c;`
- Top Poster Query
 - `select p.email_address, count(*) as c from messages as
m, messages_people as p where
m.message_id=p.message_ID group by p.email_address
order by c;`

Mailing Lists: Top Content Result (graphed)



What are people talking about?

Mailing Lists: Top Posters Result

caolanm@redhat.com	2222
bugzilla-daemon@freedesktop.org	1929
michael.meeks@suse.com	1521
michael.meeks@novell.com	1434
sbergman@redhat.com	1224
gerrit@gerrit.libreoffice.org	1076
pmladek@suse.cz	972
nthiebaud@gmail.com	942
thb@documentfoundation.org	880
kendy@suse.cz	776
bjoern.michaelsen@canonical.com	723
mstahl@redhat.com	689
serval2412@yahoo.fr	638
kyoshida@novell.com	612
tml@iki.fi	556
dtardon@redhat.com	491
l.lunak@suse.cz	490
oolst@nouenoff.nl	474
timar74@gmail.com	473

Who contributes to ML discussions?

IRC: Multiple Analysis Tools Depending on Log Format

Generates IRC stats for active people, by hour of the day, by day, most used words, quotes and more.

<http://royale.zerezo.com/irssistats/>

<http://pisg.sourceforge.net/>

<http://code.google.com/p/superseriousstats/>

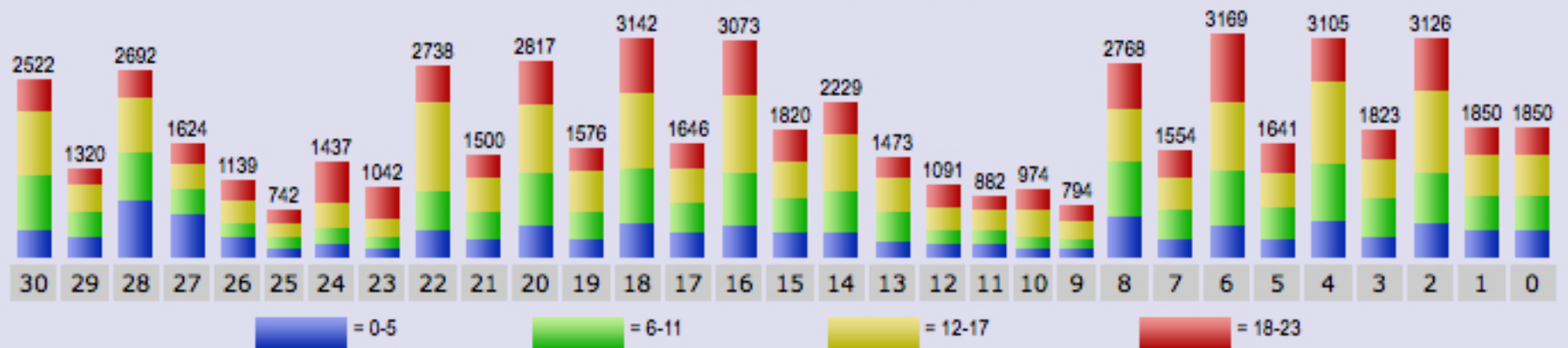
Some big numbers...		
nick	numbers	lines
alterego	is often happy :)	110
_ Nix _	is often sad :(6
gabrbedd	yells a lot !	56
lcuk	asks a lot of questions ?	312
ChanServ	often changes the modes	12
Stskeeps	likes to /kick	1
MrTroll	is often kicked	1
lcuk	posts many URLs	33
stefanoP	doesn't know wether to stay or quit	419
kimitake	often changes his nick	40
lcuk	speaks a lot of monologues	27

#puppet @ Freenode stats by MAINTAINER

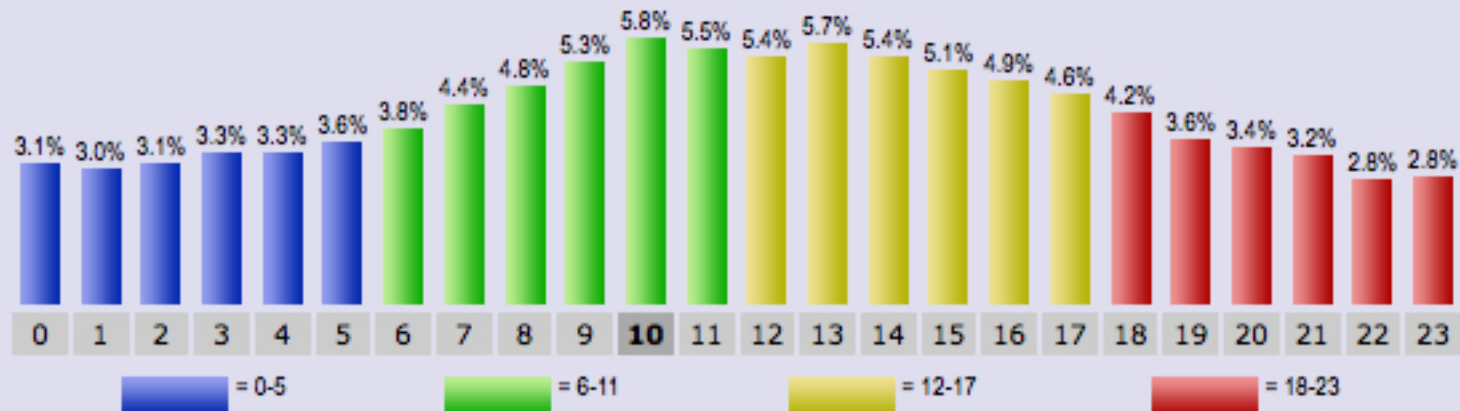
Statistics generated on Friday 12 October 2012 - 0:05:37

During this 3541-day reporting period, a total of **10243** different nicks were represented on #puppet.

Daily activity (last 31 days)



Most active times



Most active nicks

	Nick	Number of lines	When?	Last seen	Random quote
1	Volcane	86512		today	"fubada: a varaible in scope means a puppet variable"
2	jamesturnbull	25676		4 days ago	"Montreal is on the list for next year"
3	joe-mac	22564		3 days ago	"sigh, how is it possible that moving from local stored configs..."
4	fujin	21822		1771 days ago	"oh, working for Opscode on chef ;0"
5	ashp	21176		2 days ago	"lavaman: hahaha, subnets in multiple locations :("
6	lak	19439		14 days ago	"windowsrefund: if that's the case, your best bet is probably t..."
7	ohadlevy	14480		4 days ago	"slacker87: make sure you are using the right database.conf fil..."
8	whack	10360		385 days ago	"instead of file { "/etc/cron.daily/sysstat": ... }"
9	agaffney	9992		2 days ago	"fubada: cobbler has a XMLRPC API, so yes"
10	henk	9913		2 days ago	"mادتux: I already told you about apt-dater."
11	nigelk	9075		today	"aussielunix: you were great"
12	gepetto	8078		2 days ago	"axisys: no help for topic responsefile"
13	_rc	7789		2 days ago	"your sole error - 'Could not find class autof::autofs'"
14	zipkid	7775		8 days ago	"nyxynyx: actually tese q's belong in #vagrant"
15	holoway	7615		2711 days ago	"fujin: congrats on the macbook!"

Big numbers

Is **zykes**- stupid or just asking too many questions? 48.2% lines contained a question!

d3c didn't know that much either. 45.6% of his/her lines were questions.

The loudest one was **itguru**, who yelled 17.3% of the time!

Another *old yeller* was **ryancoleman**, who shouted 12.8% of the time!

It seems that **rtyler**'s shift-key is hanging: 2.4% of the time he/she wrote UPPERCASE.

For example, like this:

```
<rtyler> OH FOR SHAME
```

randybias just forgot to deactivate his/her Caps-Lock. He/She wrote UPPERCASE 2.4% of the time.

CoM is a very aggressive person. He/She attacked others **436** times.

For example, like this:

```
* cOM Slaps AdrianBroher around a bit with a large trout.</td></tr>
```

on can't control his/her aggressions, either. He/She picked on others **406** times.

Poor **gepetto**, nobody likes him/her. He/She was attacked **34** times.

For example, like this:

```
* fujin smacks gepetto
```

head seems to be unliked too. He/She got beaten **22** times.

sheldonh brings happiness to the world. 32.7% lines contained smiling faces. :)

matti isn't a sad person either, smiling 32.0% of the time.

Ned_ seems to be sad at the moment: 6.7% lines contained sad faces. :(

kesor is also a sad person, crying 6.6% of the time.

gepetto wrote the longest lines, averaging 140.6 letters per line.

#puppet average was 60.0 letters per line.

ausser wrote the shortest lines, averaging 27.7 characters per line.

[Mike] was tight-lipped, too, averaging 29.0 characters.

Volcane spoke a total of 880253 words!

Volcane's faithful follower, **ashp**, didn't speak so much: 255580 words.

Theyellowdart wrote an average of 100.00 words per line.

Channel average was 10.61 words per line.

Code Contributions: gitdm

Name	Last affiliation	Activity Start	Activity End	Commits
Caolán McNamara	RedHat	2009-10-14	2012-10-06	6961
Tor Lillqvist	SUSE	2010-09-14	2012-10-08	3252
Kohei Yoshida	SUSE	2010-05-18	2012-09-28	2351
David Tardon	RedHat	2010-07-21	2012-10-07	1727
sb	RedHat	2009-11-13	2012-10-05	1491
Michael Stahl	RedHat	2010-04-20	2012-10-06	1379
Miklos Vajna	SUSE	2010-07-29	2012-10-05	1336
Norbert Thiebaud	New Contributors	2010-09-29	2012-10-08	1234
Thomas Arnhold	New Contributors	2011-01-16	2012-08-20	1208
Michael Meeks	SUSE	2009-11-17	2012-10-04	1054
Markus Mohrhard	New Contributors	2011-03-17	2012-10-06	1045
Bjoern Michaelsen	Canonical	2010-04-19	2012-09-17	911
Fridrich Štrba	SUSE	2010-09-14	2012-10-08	897
Andras Timar	SUSE	2010-10-01	2012-10-05	778
Luboš Luňák	SUSE	2010-09-21	2012-10-04	773
Frank Schoenheit [fs]	Oracle	2010-08-13	2012-07-17	726
Joseph Powers	New Contributors	2010-10-04	2011-11-22	726
Matus Kukan	New Contributors	2011-04-06	2012-10-02	690
Cédric Bosdonnat	SUSE	2010-09-13	2012-10-03	667
Jan Holesovsky	SUSE	2009-07-07	2012-10-03	588
Rafael Dominguez	New Contributors	2011-02-13	2012-08-24	560
julien	New Contributors	2010-11-04	2012-10-07	550
Petr Mladek	SUSE	2010-05-20	2012-09-10	548
Noel Grandin	New Contributors	2011-12-12	2012-10-03	505

<http://cgit.freedesktop.org/libreoffice/contrib/gitdm-config/>

Thanks to Michael Meeks for this data

Code Contributions: Ohloh



Follow @Ohloh Sign In Join Now

Projects People Tools Meta

Projects Search...



LibreOffice

Settings Report Duplicate

370 I Use This!

Analyzed 4 days ago based on code collected 4 days ago.

Contributors : Listing

Showing 1 - 20 of 929

Filter on:



Sort by: 12 Month Commits

Name	Kudos	12 Month Commits	All Time Commits	5 Year Trend	Primary Language	First Commit	Last Commit
Caolán McNamara		2200	8812		C++	about 12 years ago	4 days ago
Tor Lillqvist		1370	4624		Make	almost 7 years ago	4 days ago
Stephan Bergmann		1130	3316		C++	about 12 years ago	7 days ago
David Tardon (Developer)		1055	1913		Make	almost 3 years ago	5 days ago
Kohei Yoshida		1048	4399		C++	over 6 years ago	14 days ago
Michael Stahl		1001	1858		C++	over 4 years ago	4 days ago
Markus Mohrhard		833	1059		C++	over 1 year ago	7 days ago
Miklos Vajna		636	1927		C++	over 6 years ago	4 days ago
Andras Timar		602	1257		XML	over 8 years ago	5 days ago
mmeeks		528	3588		C++	about 9 years ago	5 days ago

Activity

30 Day Summary

Sep 8 - Oct 8

1502 Commits
73 Contributors
including 12 new contributors

12 Month Summary

Oct 8 - Oct 8

19167 Commits
Down -9736 (-31.892%) from
previous 12 months
332 Contributors
Down -38 (-10.674%) from
previous 12 months

Community

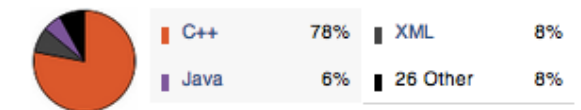
Ratings

105 users rate this project:
★★★★★ 4.4/5.0

Click to add your rating
☆☆☆☆☆

Review this Project!

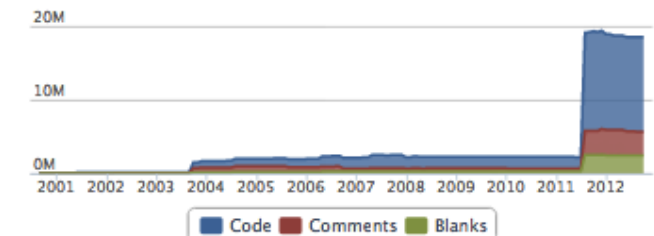
Languages



Most Recent Contributors

- Tor Lillqvist
- Fridrich Štrba
- mmeeks
- David Ostrovsky
- Norbert Thiebaud
- Tomaž Vajngerl

Lines of Code



<http://www.ohloh.net/p/libreoffice/>

New Contributors: Ohloh



Follow @Ohloh

Sign In

Join Now

Projects

People

Tools

Meta

Projects Search...



LibreOffice

Settings | Report Duplicate

370

I Use This!

Analyzed 9 days ago based on code collected 9 days ago.

Contributors : Listing

Showing 1 - 20 of 929

Filter on:



Sort by: Newest Contributors



Name	Kudos	12 Month Commits	All Time Commits	5 Year Trend	Primary Language	First Commit	Last Commit
Neven Ćosić	1	1	1	5 year commit count	XML	9 days ago	9 days ago
jallletc36	4	5	5	5 year commit count	C++	11 days ago	10 days ago
Jack Leigh	1	1	1	5 year commit count	C++	14 days ago	14 days ago
Katarina Behrens	1	1	1	5 year commit count	C++	18 days ago	18 days ago
José Guilherme Vanz	1	2	2	5 year commit count	Java	21 days ago	16 days ago
Marcos Paulo de Souza	4	5	5	5 year commit count	C++	21 days ago	15 days ago
Sergey Farbotka	1	2	2	5 year commit count	C++	26 days ago	15 days ago

<http://www.ohloh.net/p/libreoffice/>

Recent Commits with Filter: Ohloh



Follow @Ohloh

Sign In

Join Now

Projects

People

Tools

Meta

Projects Search...



LibreOffice

Settings | Report Duplicate

370

I Use This!

Analyzed 9 days ago based on code collected 9 days ago.

Commits : Listings

Showing 1 - 20 of 367

Filter on: docx




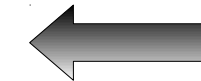
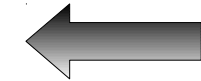
Commit Message	Contributor	Files Modified	Lines Added	Lines Removed	Code Location	Date
DocxAttributeOutput: drop unused includes	 Miklos Vajna	More...				9 days ago
rhbz#820283, fdo#55462: sw: better fix for DOCX table import crash:	 Michael Stahl	More...				13 days ago
fdo#44278: fix leaked image when import docx	 Cédric Bosdonnat	More...				14 days ago
n#782345 import DOCX w:splitPgBreakAndParaMark compat option	 Miklos Vajna	More...				16 days ago
n#782061 DOCX import: w:position should respect w:sz in w:rPrDefault	 Miklos Vajna	More...				19 days ago
docx export: always write tblInd attribute (even when equal to 0)	 Pierre-Eric Pelloux-Prayer	More...				19 days ago
docx export: add unit test for 9751056	 Pierre-Eric Pelloux-Prayer	More...				20 days ago

<http://www.ohloh.net/p/libreoffice/>

- New bugs vs. resolved bugs
 - Can't just look at monthly trends
 - Need to take release cycle into account
 - Before release: more resolved bugs
 - After release: more new bugs
- Participants
 - People who file new bugs
 - Participate in bugs (comment, etc.)
 - Careful with people who resolve bugs (usually QA)



Page statistics	
Content pages	861
Pages (All pages in the wiki, including talk pages, redirects, etc.)	3,539
Uploaded files	1,201
Edit statistics	
Page edits since MeeGo wiki was set up	41,457
Average edits per page	11.71
Job queue  length	0
User statistics	
Registered users	8,288
Active users (list of members) (Users who have performed an action in the last 91 days)	359
Bots (list of members)	0
Administrators (list of members)	10
Bureaucrats (list of members)	6
View statistics	
Views total	4,612,077
Views per edit	111.25
Most viewed pages	



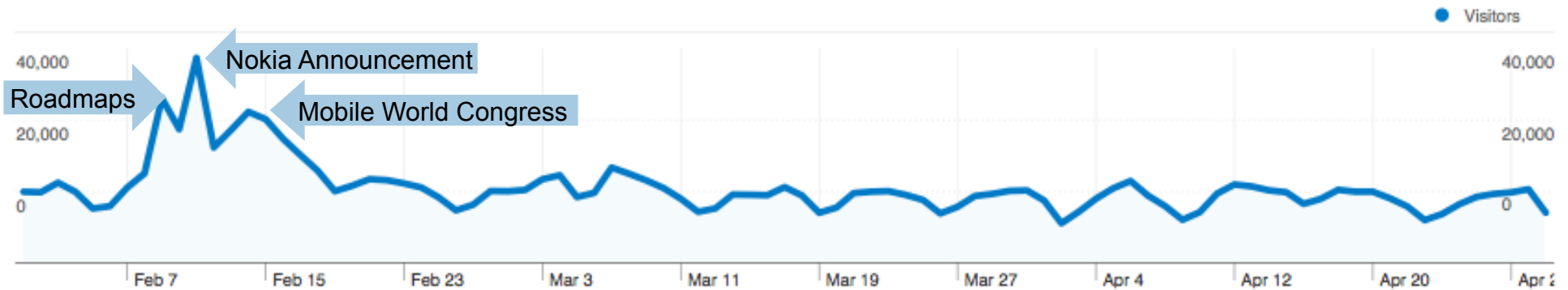
- Get Statistics
 - <http://wiki.meego.com/Special:Statistics>
 - `wget "http://wiki.meego.com/api.php?action=query&meta=siteinfo&siprop=statistics&format=yamlfm"`



Website: Google Analytics

MeeGo All Visitors Overview


Feb 1, 2011 - Apr 30, 2011
Comparing to: Site



1,022,898 people visited this site

 **1,469,962** Visits

 **1,022,898** Absolute Unique Visitors

 **4,475,647** Pageviews

Source: Google Analytics
(excludes wiki prior to Dec 21)

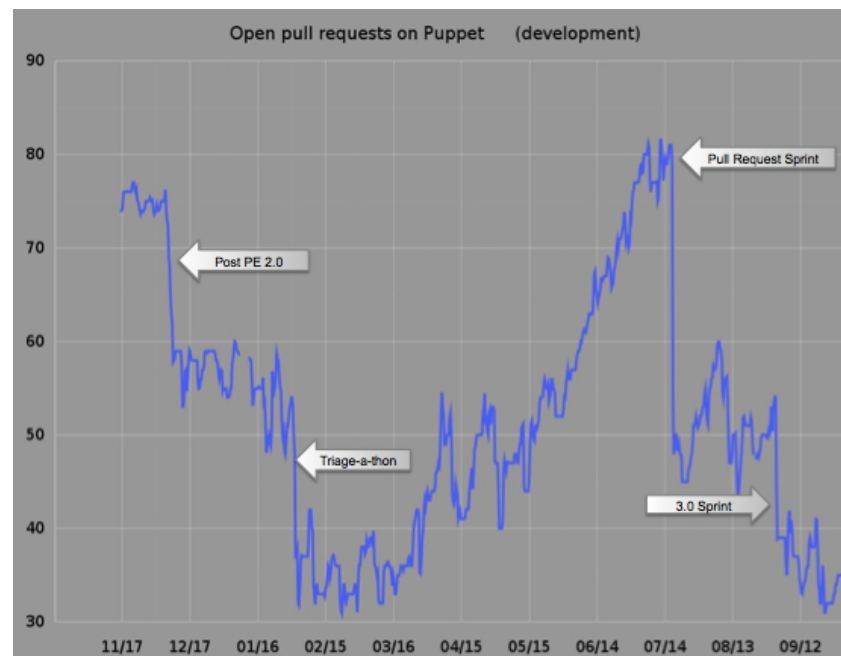
Automate: Scripts & Business Intelligence

- A less than elegant method (figure out what you want)
 - Giant bash script
 - Uses wget, awk, mysql queries, etc.
 - Dumps a bunch of csv files on my hard drive
- A better dashboard approach (WIP)
 - Open source metrics dashboard (business intelligence)
 - Uses Pentaho for reporting, runs regularly and produces a dashboard anyone can view at any time
 - <http://www.webdetails.pt/>

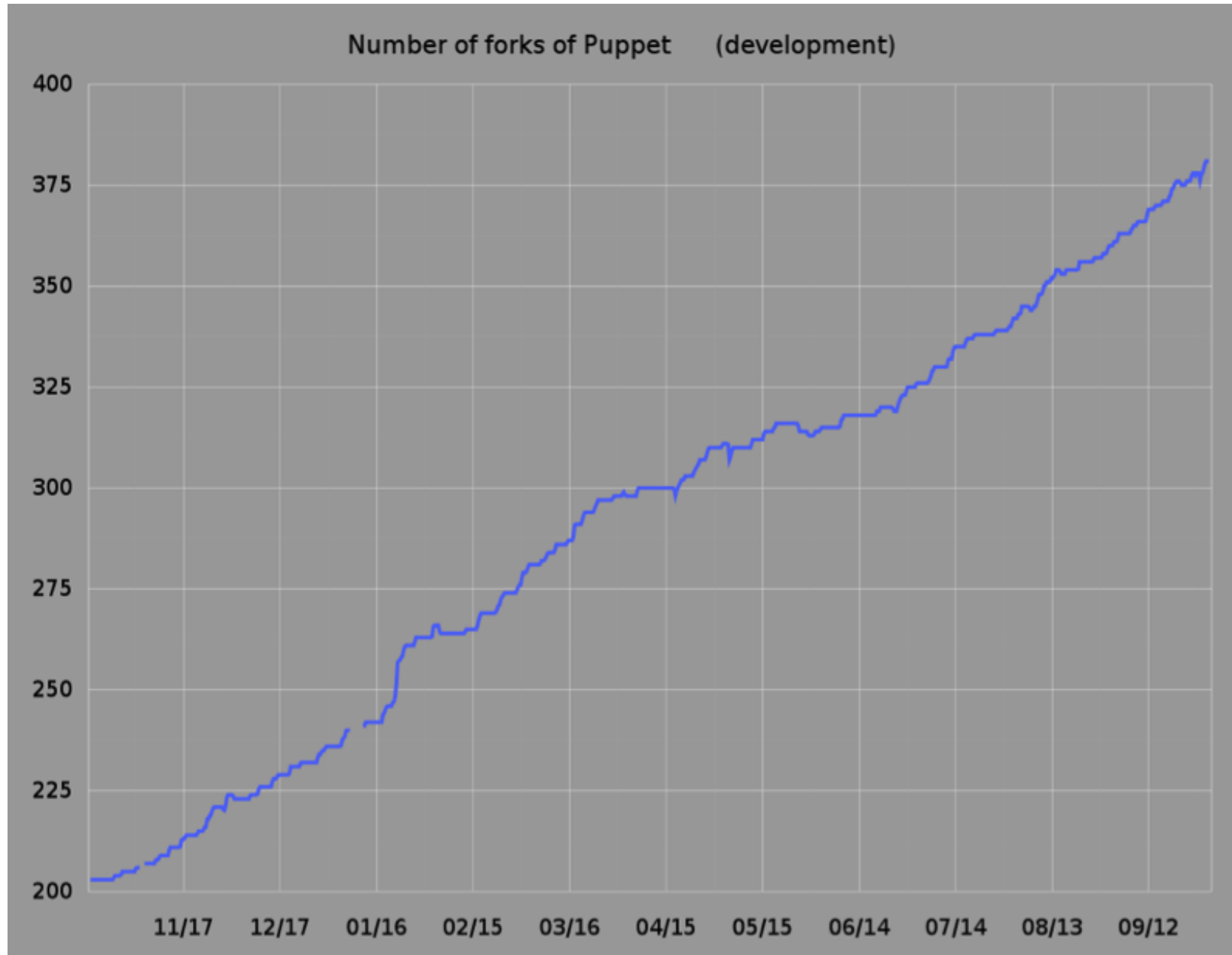


Automate: Custom Code + Graphite

- What we have now at Puppet
- Lots of custom Ruby code
- 2 Projects:
 - Gather: Gathers data from various sources
 - Metrics: Calls Gather and displays output using Graphite
- Generates one image file per graph plus a yaml file with min / max metadata
- Uses Graphite Render URL API to display data
 - http://graphite.readthedocs.org/en/0.9.10/render_api.html
- I manually add arrows to explain bumps in graphs



Example: Custom Code + Graphite



Now What?

- Report Regularly
 - Monthly – may be too often
 - Quarterly? Yearly?
- Share the reports with the community
 - <http://puppetlabs.com/community/metrics>
 - <http://wiki.meego.com/Metrics>
 - <http://tizen.org/community/metrics>
- Recognize
 - Recognize your top contributors



<http://www.flickr.com/photos/play4smee/2439494411/>

Dawn Foster
Community Manager at Puppet Labs



@geekygirl dawn

dawn@puppetlabs.com

