Interaction Networks in ML: a socio-meter for Debian

Alexandre Delanoë
Bernard Conein

Mini DebConf, 25 November 2012

Subject: "Making debian-* discussions more viable"

Russ Allbery changed the topic "From exim to postfix"

- DPL (Zack, 3 May 2012):
  - "don't feed the troll"
  - "don't feed the troll" + report abuses to listmasters and act accordingly
  - public, but contributors-only list
- A lot Energy to know what you have to ignore (or to create new rules)

Proposition: from antispam to antitroll

Existing solutions (mutt)
- ThreadIgnore.pl pipe
- Thread killer: fgrep Archive >/dev/null

Socio-meter as an Agnostic Thread Index
- Socio-meter to specify interactions in ML
- Anonymous statistics (!= teammetrics)
- Each one could filter threads (with procmail and notmuch)

Summary
- Mailing list landscape
- Threads as a group
- Threads as a network
Debian Users: Membership/Posts

(a) Nb of mails and membership/year

Paradox
- Membership increase
- Post decrease

First Corpus, May 2011

Debian, mainly from 1994
- user (646,005 mails, 17 years, 104 mails/day average)
- user-french (178,844 mails, 10 years, 49 mails/day a.)
- devel (305,106 mails, 17 years, 49 mails/day a.)
- vote (17,266 mails, 12 years, 4 mails/day a.)
- mentors (58,778 mails, 13 years, 12 mails/day a.)

Ubuntu, mainly from 2004
- user (246,233 mails, 7 years, 96 mails/day a.)
- fr (37,073 mails, 7 years, 15 mails/day a.)
- devel (33,134 mails, 7 years, 13 mails/day a.)
- devel-discuss (12,676 mails, 5 years, 7 mails/day a.)

The first corpus

Wget pages archives
for distribution in Debian Ubuntu; do
  for ml in users devel vote etc.; do
    wget $ml;
  done²

- www.debian.org/MailingLists/html
- https://lists.ubuntu.com/mbox format

But

(b) Archives and reality
Second corpus

- NNTPSERVER = 'news.gmane.org' && export NNTPSERVER
- nntp-pull gmane.linux.debian.user

Mutt threads reader

- Mutt groups threads with Subject *AND* References headers
- "/.muttrc : macro index,pager <f8> "<tag-thread><tag-prefix-cond><pipe-entry>cat /tmp/threadXXXX<enter><untag-pattern> T n<pipe-entry> /.mutt/threads.sh<enter>"
- for thread in {1..$nbThread} ; do xdoto ol F8 ; sleep 3 ; done

Debian and Ubuntu Users

- gmane.linux.debian.user, 423 275 mails
- gmane.linux.debian.user.french, 174 026 mails
- gmane.linux.debian.devel.general, 166 609 mails
- gmane.linux.debian.devel.legal, 33 617 mails
- gmane.linux.debian.devel.mentors, 54 178 mails
- gmane.linux.debian.devel.policy, 14 868 mails
- gmane.linux.debian.devel.project, 18 769 mails
- gmane.linux.debian.devel.vote, 15 552 mails

Debian and Ubuntu Users

- Increase of mail’s numbers
- Decrease of mail’s numbers after Ubuntu’s birth?
- But the cause of the phenomenon is not clear
- It cannot be explained only by Ubuntu itself
- Other factors should be taken into account:
  ⇒ Less bugs, less requests, other tools, other areas for posting
Decrease of thread's number?
No clear explanation
Less topics to be discussed
But does it imply that there are less discussions?
NO

Increase of mails number:
Continuous for Debian user
Discontinuous for Debian Devel
Size of thread counts?
Maybe less topics but complex topics
Other explanation: the size is not enough as the thread structure might count.
Joey Hess’ Stuff, 30 October 2005: it’s all about pattern recognition

**References**

- Joey Hess: it’s all about pattern recognition
  - Two types of thread graph
    - Star shaped = responses as answers
    - Line shaped = responses as oriented to the prior response
  - Examples
    - (o) Start shaped thread: vote by agreement
    - (p) Line shaped thread: vote by deed

Joey Hess explains how to subscribe to mailing lists with a combined total posts of 2000 or more per day, and live. It’s all about pattern recognition.

**Private**

- Worthwhile
- Too late

**Take it**

- Interesting
- Obvious Answer

**Think before!**

- Foo?
**Shape (depth) of threads**

- (q) Average depth by year

<table>
<thead>
<tr>
<th>lidi</th>
<th>debian-devel</th>
<th>debian-mentors</th>
<th>debian-user</th>
<th>debian-user-french</th>
<th>debian-vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>dd</td>
<td>100</td>
<td>25</td>
<td>11</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>dm</td>
<td>25</td>
<td>100</td>
<td>10</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>du</td>
<td>11</td>
<td>10</td>
<td>100</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>duf</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>dv</td>
<td>36</td>
<td>19</td>
<td>5</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>

(s) Top 100 contributors contribution each list

- Contributors of each list are "specialized"
- Entry barriers

**Debian and Ubuntu passive membership and posts**

- (t) % of mails and passive members/year

- Members who stop posting after posting one at least
- Only active members re-intervene
### Calcul

\[
ThreadIndex = \frac{\sum_{1}^{N} Mails_i}{\sum_{1}^{N} References_i + 1}
\]

### Structure Thread index

- **Index comprehension:**
  - Strictly Star Shape thread Index = 1
  - Strictly Line Shape thread Index tends to 0

- **Comprehensive analysis**
  - Star Shape = vote (secunded)
  - Strict line shape = collaboration
  - Big Mix analysis of flaming

---

### Consider thread as a directed graph

#### Diachronic Thread index View

<table>
<thead>
<tr>
<th>Years</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>0.0</td>
</tr>
<tr>
<td>2004</td>
<td>0.1</td>
</tr>
<tr>
<td>2006</td>
<td>0.2</td>
</tr>
<tr>
<td>2008</td>
<td>0.3</td>
</tr>
<tr>
<td>2010</td>
<td>0.4</td>
</tr>
<tr>
<td>2012</td>
<td>0.5</td>
</tr>
</tbody>
</table>

- **Index evolution through time**
  - Index evolution over time: 4/4 (4/4)
How many authors per thread?

- Number of dyads increases whereas others are stable (Monad, triad, tetrad, pindad and more)
- Relation between number of authors and structure of the thread [Concin and Latapy, 2008]

Content analysis

- Figure: Factor analysis, Subject/From

Stability of the Thread Index

- Figure: Factor analysis, Subject/Index
Stability of the Thread Index

Figure: Factor analysis, From/Index

Interview 2010

Raphaël Hertzog
You contribute regularly to Debian mailing lists, yet I don’t remember any aggressive/frustrated mail of you. How do you manage that? Are you avoiding heated discussions?

Joey Hess
Rats, sounds like all my wonderful flames of years past have been forgotten! Seriously though, after I noticed thread patterns, I started trying to avoid participating in the bad patterns myself. Now I limit myself to one expression of an opinion, and I don’t care who gets the last word. If people can’t be convinced, it’s time to find another approach to the problem. Also, code talks.

Source

Conclusion

Work in progress
- To analyze mailing-list archive can be useful
- To produce not only a new meter
- But an agnostic thread index
- That everyone could manage with (ham/troll)
- But it supposes that everyone have access
- to headers of the mailing-list archive
→ How to preserve privacy and ignore trolls?

References

Questions?

conein@unice.fr | debian@delano.e.org

2012, Paris - France